Methods for Identification and Quantitation of Small Molecules and Peptides in Bioenergy, Metabolism, and Proteome Research

By

Arne Ulbrich

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: 12/09/2015

The dissertation is approved by the following members of the Final Oral Committee:

Joshua J. Coon, Professor, Chemistry
Lloyd M. Smith, Professor, Chemistry
Lingjun Li, Professor, Pharmacy
Shannon S. Stahl, Professor, Chemistry
David J. Pagliarini, Associate Professor, Biochemistry

METHODS FOR IDENTIFICATION AND QUANTITATION OF SMALL MOLECULES AND PEPTIDES IN BIOENERGY, METABOLISM, AND PROTEOME RESEARCH

Arne Ulbrich

Under the supervision of Professor Joshua J. Coon

At the University of Wisconsin-Madison

Abstract

This dissertation considers the development of methods and hardware for identification and quantitation of small molecule and peptide analytes. Specifically, (1) LC-MS, GC-MS, and LC-UV techniques for the identification and quantitation of lignin-derived aromatics and the application of those techniques, (2) LC-MS analyses of low-abundance lipids with special emphasis on Coenzyme Q metabolites, (3) the development of two NeuCode quantitation methods, (4) methods for the improvement of precursor ion selection in proteomic experiments, and (5) the improvement of proteome digest separations by custom-designed and manufactured column thermostats are explained in detail in the following chapters.

Chapter 1 provides a brief overview and some background on the measurement of molecules by chromatography coupled to either a mass spectrometer or a spectrophotometer. The factors that dictate the choice of technique are discussed, as are determinants of measurement precision and accuracy. Special attention is given to factors relevant to the research projects discussed in this dissertation. **Chapters 2** and **3** discuss the development and assessment of three novel NeuCode quantitation methods for organic acids and peptide digests by LC-MS and GC-MS. **Chapter 4** describes segmenting the mass spectrometer scan range for optimization of precursor selection in proteomic studies. **Chapter 5** reports on the design, manufacture, and

characterization of custom column ovens for improved chromatographic protein digest separations. **Chapter 6** describes the applications of LC-MS, LC-UV, and GC-MS methods I developed to the depolymerization of lignin to value-added aromatics, the quantification of fermentation inhibitors during bioethanol production, and the characterization of proteins involved in Coenzyme Q biosynthesis.

Acknowledgements

I thank Professor Coon for his guidance during my time in his research group. He has provided good research opportunities, ensured the availability of excellent resources, and recruited talented fellow graduate students, post-doctoral researchers, and staff scientists. Of course, the work presented here is not mine alone but its success heavily depended on the collaboration and advice from past and present Coon Group members as well as many outside collaborators and past mentors. I would like to single out Aaron Ledvina, Michael Westphall, and Derek Bailey for their extensive mentoring and help especially at the beginning of my graduate career. Furthermore, I acknowledge the long and productive collaborations with members of the Pagliarini and Stahl Research Groups and the Great Lakes Bioenergy Research Center (GLBRC), especially Jon Stefely, Danielle Lohman, and Ali Rahimi. Drs. Sipe and Porterfield at Hampden-Sydney College have encouraged me to major in chemistry and have been excellent mentors and teachers. Without Dr. Sipe's encouragement, who also received his PhD from the University of Wisconsin, I would not have applied to attend graduate school in Madison. I would also like to thank Jae Schwartz, John Syka, and Jens Griep-Raming for facilitating two productive and enlightening industrial internships at Thermo Fisher Scientific in San Jose, CA and Bremen, Germany, my hometown. I thank my parents and my uncle Horst and aunt Hiltrud for supporting me and encouraging me in my studies. Rebecca has patiently accepted my long work hours and always supported me over the past 1.5 years. Finally, I would like to thank the

members of my committee: Professors David Pagliarini, Lingjun Li, Lloyd Smith, and Shannon Stahl.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

LC-MS ..... liquid chromatography – mass spectrometry

GC-MS..... gas chromatography – mass spectrometry

LC-UC..... liquid chromatography – ultraviolet absorbance

NeuCode..... neutron encoded quantitation

Q..... Coenzyme Q

CoQ ..... Coenzyme Q

NMR ..... nuclear magnetic resonance

IR..... infrared

UV/Vis..... ultraviolet/visible light

ppb..... parts per billion

nm..... nanometer

log P ..... octanol/water partition coefficient

GC ..... gas chromatography

Th ..... Thomson

ESI..... electrospray ionization

Da..... Dalton

ppm..... parts per million

MS ..... mass spectrometry

%T ..... percent transmittance

ν ..... nu, frequency

ε..... epsilon, extinction coefficient

FAME..... fatty acid methyl ester

mDa..... Millidalton

LysC ..... Endoproteinase Lys-C

FDR ..... false discovery rate

M..... molarity (mol/L)

FT-ICR..... Fourier transform ion cyclotron resonance

R..... arginine

H ..... histidine

Lys ..... lysine

FT-MS ..... Fourier transform – mass spectrometry

m/z ..... mass-to-charge

R ..... resolution

MS^1 ..... survey scan mass spectrometry

MS/MS ..... tandem mass spectrometry

SILAC ..... stable isotope labeling in cell culture

HPLC ..... high performance liquid chromatography

nHPLC ..... nano-HPLC

TOF ..... time of flight

$[M+H]^+$ ..... protonated molecular cation

µL ..... microliter

µm ..... micrometer

C18 ..... octadecyl

ng ..... nanogram

eV ..... electron volt

GCxGC ..... 2-dimensional GC

NaOH ..... sodium hydroxide

RP ..... resolving power

HCD ..... higher energy collisional dissociation

XIC ..... extracted ion chromatogram

UHPLC ..... ultra-high performance liquid chromatography

σ ..... standard deviation

fg ..... femtogram

pg ..... picogram

SIM ..... selected ion monitoring

AOCS ..... American Oil Chemists' Society

KOH ..... potassium hydroxide

mTh ..... Millithomson

GPF ..... gas phase fractionation

DDA ..... data dependent acquisition

g ..... gravity

EDTA ..... ethylene diamine tetraacetic acid

DTT ..... dithiothreitol

Hz ..... Hertz

mM ..... millimolar

rpm ..... revolutions per minute

UPLC ..... ultra performance liquid chromatography

i.d ..... inner diameter

ACN ..... acetonitrile

MeOH ..... methanol

ITCL ..... ion trap control language

NCE ..... normalized collision energy

psi ..... pounds per square inch

nL ..... nanoliter

W ..... watt

PID ..... proportional integral derivative

RTD ..... resistance temperature device

CAD ..... computer assisted drawing

GLBRC ..... Great Lakes Bioenergy Research Center

HESI ..... heated electrospray ionization

AGC ..... automatic gain control

AFEX ..... ammonia fiber expansion

µM ..... micromolar

CID ..... collision-induced dissociation

ACSH ..... AFEX treated corn stover hydrolysate

CoA ..... coenzyme-A

PM ..... photosynthetic medium

SIS ..... Sistrom’s minimal medium

wt% ..... percent by weight

H ..... p-hydroxyphenyl

G ..... guaiacyl

S ..... syringyl

M_w ..... weight average molecular weight

M_n ..... number average molecular weight

k ..... rate constant

OD ..... optical density

PKL ..... protein kinase-like

p ..... p-value

Δ ..... knockout

Chapter 1 – Introduction

The fundamental task of analytical chemistry

In chemical research and process development it often becomes necessary to both identify and quantify individual components that are part of complex mixtures and may be present only in trace amounts. All analysis comes down to these two questions: **(1) "Which substances are present?"** and **(2) "How much of each substance is present?"** in a given sample. Neither identification nor quantitation of such compounds is trivial. Nuclear magnetic resonance spectroscopy (NMR), infrared spectroscopy (IR), ultraviolet/visible spectroscopy (UV/Vis) are powerful tools for molecular structure determination and quantitation but their usefulness to interrogate complex mixtures and compounds present at low concentrations is limited.[1,2] Mass spectrometry provides increased sensitivity and allows the analysis of mixtures of compounds but trace components might be masked by the presence of a strong matrix signal or by other analytes with higher response factors. Introducing a chromatographic separation in front of the detector allows the detector to, ideally, only encounter one species at a time, which greatly improves specificity and selectivity. The combination of a chromatograph with a mass selective detector, which is very sensitive and provides an additional dimension of separation, by molecular weight, is a very powerful one. This combination, in some cases, allows the detection of analytes at levels well below one part per billion (ppb).[3] The most common couplings are those of a liquid chromatograph to a UV/Vis detector or of a gas chromatograph to a mass spectrometer but many other combinations exist.[2,4] This thesis will predominantly discuss LC-MS methods, i.e. the coupling of a liquid chromatograph to a mass spectrometer but GC-MS and LC-UV methods were also developed for the research presented here.

Choice of Analytical Technique

The choice of analysis technique for a set of analytes in a specific type of sample depends on many factors, among them (1) availability of the instrumentation, (2) ability of the detector to detect the analyte of interest, (3) concentration of the analyte, (4) volatility of the analyte, (5) complexity of the analyte mixture, (6) ability of the technique to identify compounds, and (7) nature and complexity of the matrix. All of these factors have guided my analysis decisions during my graduate career.

Let us consider the research described in **Chapter 6**, the analysis of natural substituted aromatics. The compounds that were analyzed for one project arise from lignin, whose depolymerization yielded a number of major unique products. Anywhere from 6 – 20 products were observed, all of which contain an aromatic ring and several polar functional groups. All compounds are soluble in common organic solvents and at least moderately soluble in water. The goals of the project were to identify all major components of the mixture and to accurately quantify the conversion of lignin into monomers. Trace components were not of interest. Considering the factors outlined in the previous paragraph, I decided to use separation by reverse phase liquid chromatography coupled to both a UV/Vis and a high resolution mass spectrometer. Several properties of the analytes guided this selection of analysis system. The polar moieties on the depolymerization products, phenol groups and carboxylic acids, enable the compounds to exist in relatively stable anionic forms to allow mass spectrometric detection and all components were expected to show strong UV absorbance around 254 nm due to the presence of a six-membered aromatic ring.[5] The components were not sufficiently volatile for GC analysis without previous derivatization, which would have introduced additional variability. A chromatographic separation was necessary because the matrix contained strongly acidic catalyst species as well as any remaining polymeric lignin that could cause signal suppression were the components not separated. Reverse phase high performance liquid chromatography

(HPLC) was chosen as the separation method because all analytes were well-retained and showed good separation from one another. Structural information and elemental composition of the compounds necessary for identification was derived from LC-MS/MS, GC/MS, LC-MS, and LC-UV. The points above illustrate that at the beginning of each new research project, the analyst must make an educated decision on which technique is best suited for the project at hand and this decision will greatly influence success or failure of the project.

Compound Identification

Let us reconsider the first part of the fundamental task of the analyst: **How does one determine which substances are present in a sample?** The analyst can use several pieces of evidence arising from a compound's chromatographic behavior and its spectra. The assembled spectra and relative retention times can then either be interpreted de-novo or compared to databases.[5-7] The work presented here used gas and liquid chromatography coupled to mass spectrometers and UV detectors.

A compound's polarity can be roughly deduced from its retention time relative to known compounds. For example, in reverse phase HPLC the hydroxybenzoic acids are expected to elute earlier than benzoic acid due to their increased polarity since compounds are separated by their differential partitioning between a polar mobile phase and a non-polar stationary phase.[8,9] More polar compounds, i.e. those with a smaller *log P* value will spend a larger fraction of time in the mobile phase and will thus reach the detector earlier. In GC the situation is reversed. Benzoic acid will elute before the hydroxybenzoic acids since compounds are separated based on their volatility. The hydroxybenzoic acids have an increased ability for intermolecular hydrogen bonding, therefore they spend less time in the gas phase in the helium stream and more time condensed onto the column's stationary phase.[10] After completing their journey

through the chromatographic column compounds enter a detector, which was most often a mass spectrometer during my graduate work.

Mass spectrometry can provide information on the molecular structure of compounds and the types of functional groups present. If high resolution MS (here, an Orbitrap) is used, the elemental composition of the compound can be determined. This requires a soft ionization technique, such as electrospray ionization, that does not cause source fragmentation (**Figure 1.1 A**).[11-14] Fragmentation of molecules can also be desirable to obtain insights into atom connectivity within molecules. Fragmentation can be effected by many routes including collisions with nitrogen gas and electron bombardment.[15] Classes of molecules give rise to characteristic fragment ions after dissociation and that information plays an important part in the determination of the identity of an unknown (**Figure 1.1 B**).[7,11] For example, methyl esters and methyl amides, whose analysis is discussed in **Chapter 4**, give rise to class-identifying ions at 74 and 58 Th, respectively.[16,17] Coenzyme Q biosynthesis intermediates also give rise to characteristic head-group fragments, a property that I exploited to increase sensitivity and selectivity of an LC-MS quantitation method (**Chapter 6.2**).[18,19] Other lipid molecules reveal their fatty acid side chains upon fragmentation of their anions, rather than cations. Furthermore, the preferred formation of cations or anions gives valuable information on the nature of functional groups present in a given molecule. Electrospray ionization (ESI) allows for the formation of both anions and cations by simply reversing the polarity of the applied electric field without the need for instrument modifications. A strong negative ion mode signal obtained from an unknown suggests carboxylic acid, nitro, or phenol functionalities whereas a strong positive ion signal indicates the presence of amine, alcohol, ester or a number of other functionalities. The absence of signal in one polarity can be used to exclude possible structures. I used this approach in the identification of unexpected nitrated aromatics, as described in **Chapter 6.1**.[20]

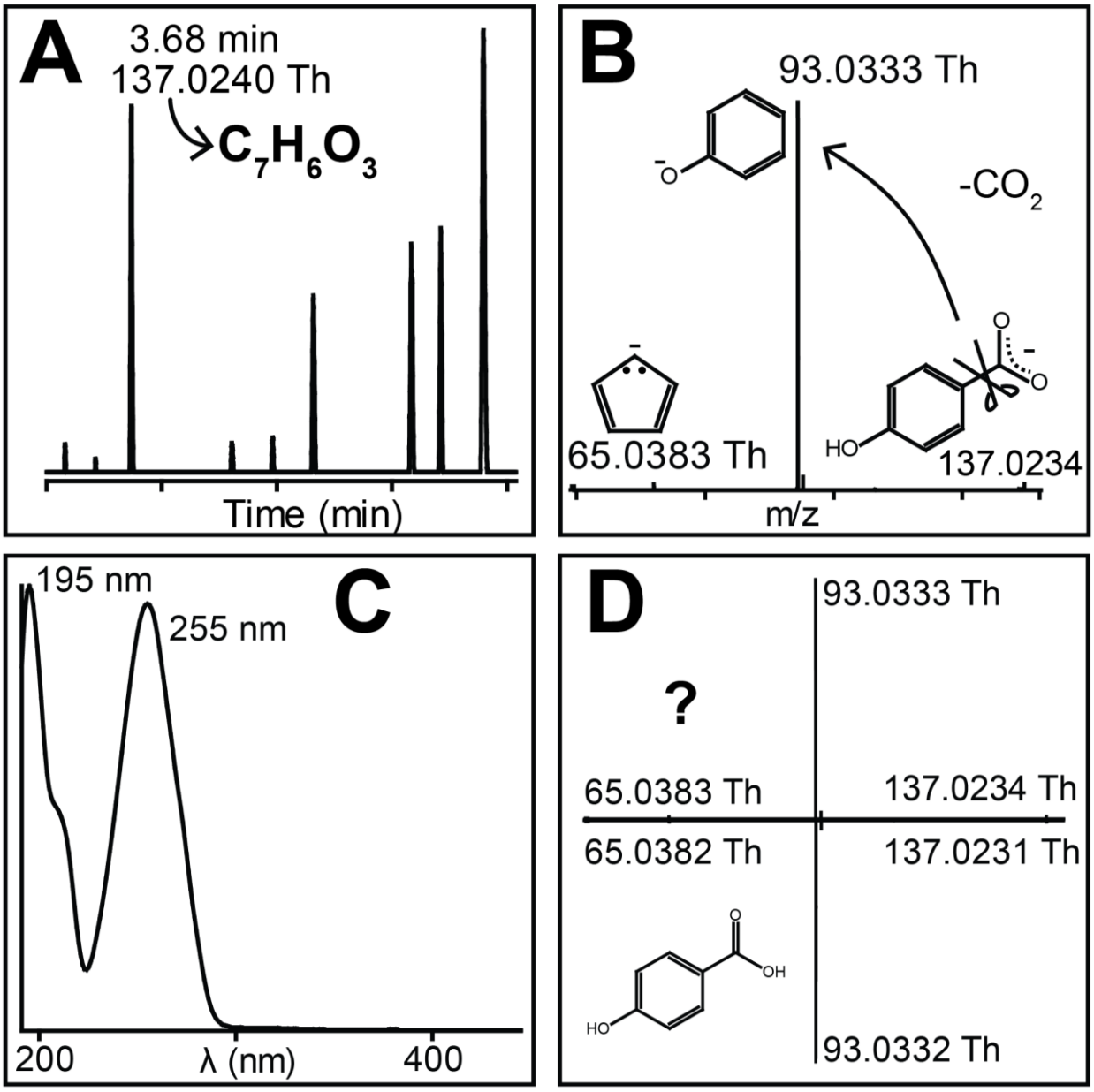


Figure 1.1 Determination of the identity of an unknown compound. **A** Determination of relative retention time and elemental composition by LC-MS. **B** Structural insights from LC-MS/MS. The neutral loss of 44 Th indicates a carboxylic acid. The anion at 93 Th has the mass of a phenolate anion and the peak at 65 Th indicates the cyclopentadienyl anion, which is commonly observed from fragmentation of aromatics. **C** The UV spectrum confirms that the compound is a substituted aromatic. Expected peaks for o-, m-, and p-hydroxybenzoic acid are at 237 nm, 237 nm, and 255 nm, respectively.[5] **D** Confirmation of the hypothesis that the unknown is p-hydroxybenzoic acid by comparison of its MS/MS spectrum to that of a standard.

Elemental composition can be directly determined from an accurate measurement of the mass-to-charge ratio of an ion due to the varying mass defects of different elements.[21] For example, acetophenone (C_8H_8O) and propylbenzene (C_9H_{12}) both have a molecular weight of 120 Da (**Figure 1.2 A**). The peaks are indistinguishable by a unit-resolution mass analyzer, such as a quadrupole (**Figure 1.2 B**). Due to oxygen's smaller mass defect the accurate monoisotopic masses of the two compounds differ by 0.036 Da with acetophenone's accurate mass being 120.0575 Da and propylbenzene's being 120.0939 Da. This phenomenon is also the foundation of the NeuCode quantitation technique pioneered by the Coon Research Group (**Chapters 2 & 3**).[22] A high mass accuracy mass analyzer will be able to assign an accurate mass and assign an unambiguous elemental composition to each measured mass peak (**Figure 1.2 B**). A mass accuracy of about 100 ppm is required to distinguish between the compounds in this example but the more atoms a molecule contains, the smaller each atom's relative contribution to the total mass defect becomes and much higher mass accuracy will be required for determination of elemental composition. In addition, the presence of nitrogen and other elements will further crowd the mass defect space requiring yet greater mass accuracy. In practice, to reliably assign elemental compositions to small molecules a mass accuracy of less than 10 ppm is needed.[23]

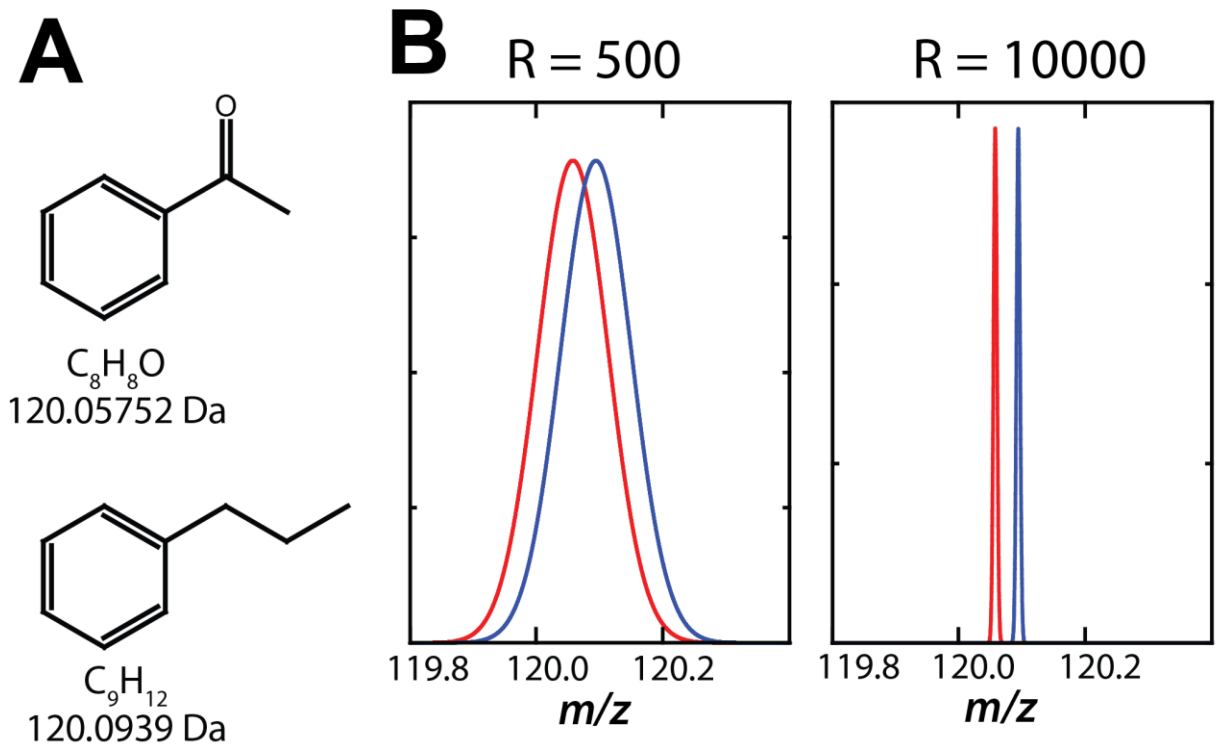


Figure 1.2 Resolution of isobaric non-isomeric compounds by high resolution accurate mass MS. **A** Acetophenone and propylbenzene both have a molecular weight of 120 Da but differ in mass defect due to differing atomic compositions. **B** The two mass peaks cannot be resolved on a unit resolution mass analyzer but two distinct peaks are observed when a high resolution mass spectrometer is used.

The characteristic absorbance of ultraviolet (UV) and visible light of a molecule can aid in its identification by detecting conjugated double bonds and certain functionalities (**Figure 1.1 C**). However, UV spectroscopy lacks the power and scope of tandem mass spectrometry and nuclear magnetic resonance spectroscopy for structure elucidation because absorption bands are much wider and fewer in number than MS or NMR peaks.[7] As does tandem mass spectrometry, UV spectroscopy exploits differences in bonding and connectivity of atoms inside a molecule. Different molecules absorb light at different discrete wavelengths and thus give a characteristic and unique spectrum or fingerprint. A valence electron is promoted to a higher energy level only by absorption of certain wavelengths of light that match the gap between energy levels. The size of this gap depends on a host of factors, such as conjugation and substituents on aromatic rings.[5,24] This absorption is what UV spectroscopy measures. The measurement is expressed as a plot of absorbance versus wavelength of light. Absorbance is defined as:

$$A = -log_{10}(T) \quad eq.1.1$$

T is transmittance, the fraction of light transmitted by the solution. The advantage of plotting absorbance rather than transmittance is that wavelength regions where a compound absorbs light are represented by positive peaks unlike the convention to show IR spectra as %T versus $1/\nu$.

UV spectroscopy's power to elucidate molecular structure is limited but nevertheless important evidence can be obtained from UV/Vis spectra. Key evidence in the identification of unexpected nitro compounds in one of the papers that are part of this thesis was obtained by UV/Vis spectroscopy (**Chapter 6.1.3**).[20] Important UV chromophores are relatable to structural features. Alkanes, non-conjugated alkenes, alkynes, and most other functional groups absorb only below 200 nm or only very weakly above 200 nm and these peaks are generally masked by the LC

solvent (water, methanol, or acetonitrile) unless a GC-UV is used, which is rare.[25] However, the absence of absorbance bands above 200 nm can help characterize a compound by exclusion. A notable exclusion to the non-absorption of most functionalities is nitro groups, which absorb above 250 nm and out to >300 nm. UV spectroscopy is most useful for the characterization of conjugated polyenes. The most important class of conjugated polyenes are compounds containing aromatic rings.[5] Most small molecules investigated for the projects described in this thesis contain this functionality. Aromatic compounds show strong absorbance (ε = 2000 – 10000) for the $\pi \rightarrow \pi^*$ transition at 180 – 230 nm. When a substituent is present a second band around 240 nm appears that also absorbs strongly (ε = 10000 – 30000) and substituents containing lone pairs further shift the absorbance to longer wavelengths as do additional substituents on the ring or polycyclic systems.[4,5,7,26] Rules exist to allow fairly accurate prediction of $\pi \rightarrow \pi^*$ absorption maxima of substituted aromatics and conjugated dienes.[5,27] Following these rules, one can predict absorption maxima to within a few nm. A calculation of the maximum of p-hydroxybenzoic acid gives 255 nm, which is identical to the experimental value (**Figure 1.3 A, 1**). Vanillic acid, a more complex molecule, gives 262 nm compared to an experimentally determined value of 260 nm (**Figure 1.3 A, 2**). This predictive power and being able to compare the unique fingerprint spectra to authentic standards makes UV spectroscopy useful, although its de-novo structure determination power is limited. The technique is very well suited for quantitative applications, however.

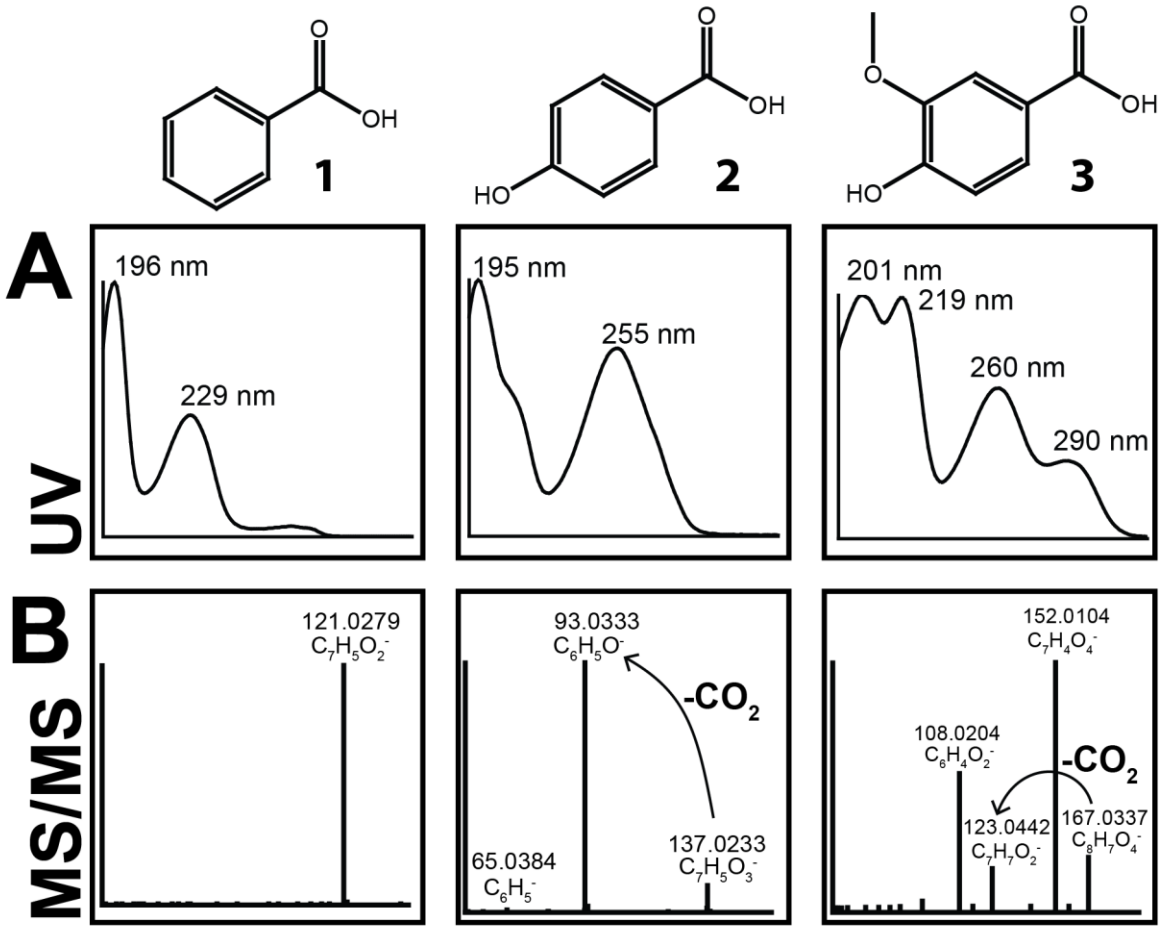


Figure 1.3 Relationships of UV spectra and MS/MS spectra of benzoic acid, 4-hydroxybenzoic acid, and vanillic acid, three similar compounds. **A** The UV absorption maxima shift to longer wavelengths when substituents are added to the aromatic ring. And, depending on the symmetry of the molecule, additional bands appear. Had another methoxy group been added in the 5 position, no additional bands would have appeared. **B** The tandem mass spectra of the compounds also increase in complexity. The neutral loss ion at -44 Th for carbon dioxide is absent in benzoic acid, since the charge will remain on the CO_2 molecule, which is below the low-mass cutoff. Vanillic acid also shows methyl loss (152 Th) and methyl + CO_2 loss (108 Th).

Compound Quantitation

The second part of the fundamental task of the analytical chemist is to determine **how much of a compound is present in a mixture**. Quantitation of compounds can either be relative, expressed as fold-changes between two samples or it can be absolute, expressed as a specific concentration of a compound in a sample. Relative quantitation was used to determine the effects of gene knockouts and single point mutations on coenzyme Q biosynthesis (**Chapter 6.2**).[18,19] Proteome remodeling in response to disease, stress, or mutations is also usually expressed as relative fold-changes in protein abundances.[16,18] It is easier to provide quantitative information for many compounds at once using relative quantitation rather than absolute quantitation since authentic standards and calibration curves are not required. In fact, not even the identity of a compound has to be known. Determination of concentration of a substance in solution or yield of a reaction by LC-MS calls for absolute quantitation, which requires the preparation of calibration curves from standards. This is more labor intensive and the number of quantified compounds is generally much lower but accurate quantitation is needed for many applications, such as the GLBRC projects highlighted in **Chapter 6.1.**[20,28,29] In brief, a serial dilution of a standard solution of an analyte of interest is prepared and the resulting samples are injected to establish a (hopefully) linear relationship between detector response and analyte concentration (**Figure 1.4**). Then a sample with an unknown concentration is analyzed and the recorded detector response can be related back to a concentration using the calibration curve that was made from the serial dilutions of the analyte standard. To ensure consistency, an internal standard that is either an isotopically labeled version of the analyte or structurally similar to the analyte can be used. It is added to all sample and standard injections at a constant concentration. All results are then normalized to this internal standard.

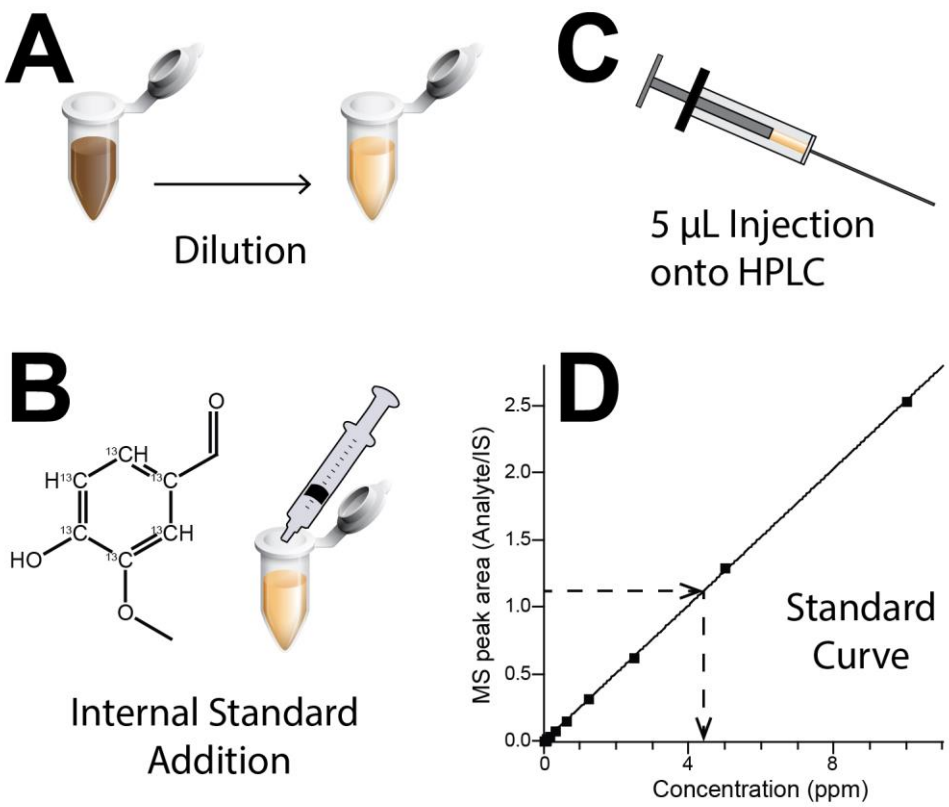


Figure 1.4 Absolute Quantitation by LC-MS. **A** Dilution of sample to a suitable concentration for LC-MS analysis (~ 10 ppm is typical). **B** Addition of an isotopically labeled internal standard to correct for effects compromising quantitative accuracy. **C** Injection of the sample – 5 µL is a typical injection volume for (U)HPLC **D** Determining concentration in a sample from a calibration curve constructed from injections of a series of dilutions of a standard.

The response factors of different aromatic compounds tend to be more alike from UV detection than MS detection and the linear dynamic range of UV detection can be larger than that of MS detection, at least with the specific instrumental setup used for much of the work described here. Sensitivity and selectivity are much lower than MS, however. This makes UV a good technique for precise quantitative analysis of relatively simple mixtures of major components, which is exactly what was needed for quantitation of lignin depolymerization products. However, UV absorption bands of different compounds more than likely overlap, which makes complete chromatographic separation of all peaks to be quantified generally necessary. It is thus much harder to achieve short run times and the analyst cannot be confident that there aren't minor components hidden under other peaks. One of the many advantages of MS detection is the lowered chromatographic requirements, which is why MS detection was used for detection of lignotoxins in fermentation media (**Chapter 6.1.2**). Proteome analysis deals with even more complex mixtures of peptides, so MS or tandem MS quantitation is used exclusively.

Analysis of complex peptide mixtures is challenging in several ways. The dynamic range of protein expression is vast, spanning more than five orders of magnitude.[30,31] Many peptides are co-eluting at each point in the LC gradient because a tryptic digest yields 100,000s of peptides while the peak capacity of even a very good separation is generally under 500 but can be increased to several thousand by using orthogonal prefractionation.[32,33] The mass spectrometer is usually operated in data-dependent mode. In this mode a full mass scan is taken to acquire intact masses of peptides eluting at each point in time. Then the 10, 15, or 20 most abundant ions are selected for fragmentation to extract sequence information and sometimes quantitative information.[34] This mode of acquisition biases identifications and quantitation towards a subset of highly abundant peptides (and thereby proteins). To limit this bias, special modifications to the acquisition algorithms have been made that allow the exclusion of isotopic peaks due to the presence of ^{13}C and that exclude peaks selected for a set amount of time after having been

fragmented.[35] Both of these features limit the redundant re-sampling of abundant peptides. To further limit the bias toward abundant species, we developed a method to divide the mass range into several sections or tiles to allow the selection of low abundance peptides. That work is described in detail in **Chapter 4**. The Coon Research Group has also pioneered a new technique for peptide quantitation using LC-MS that takes a creative novel approach to peptide quantitation. The technique exploits the small differences in mass defect between ^{13}C, ^{15}N, and ^{2}H to allow simultaneous quantitation of multiple samples in one analysis. The technique is similar to the more established use of isobaric mass tags but high resolution MS^1 scans rather than MS/MS spectra are the source of quantitative information. The technique has been established for use in proteomics and one such application is discussed in **Chapter 2**. We recently expanded the technique to small molecule analytes and GC-MS analysis. A published application to organic acid quantitation and unpublished work for multiplexing of fatty acid methyl ester (FAME) analysis are described in **Chapter 3**.

In its entirety this thesis should provide a good overview of how an analyst goes about determining which compounds are present in a sample and how much of each compound is present. Many research projects require an answer to those fundamental questions, with the work described here ranging from the improvement of bioethanol production to identifying the function of uncharacterized enzymes. The reader will hopefully gain an appreciation of the many different approaches and challenges to identification and quantitation of peptides and small molecules using hyphenated analysis techniques.

Chapter 2 – Neutron Encoded Protein Quantification by Peptide Carbamylation

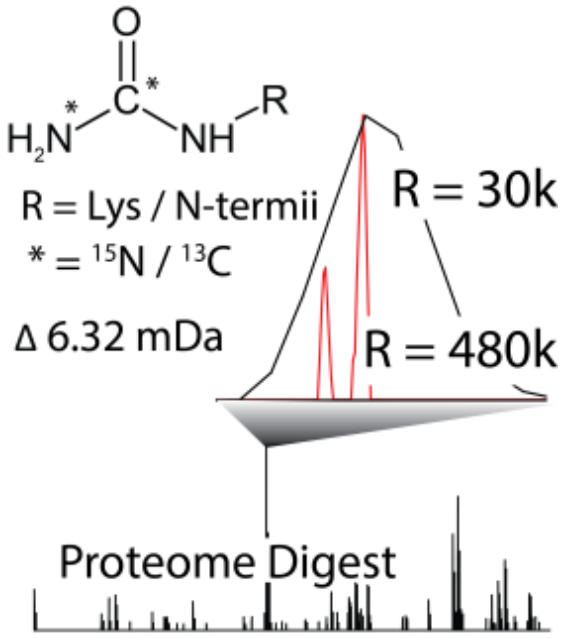
A.U. designed experiments, performed experiments, analyzed data, and wrote the paper.

This chapter has been previously published:

Ulbrich, A., Merrill, A. E., Hebert, A. S., Westphall, M. S., Keller, M. P., Attie, A. D., & Coon, J. J. (2014). Neutron-encoded protein quantification by peptide carbamylation. *Journal of the American Society for Mass Spectrometry*, *25*(1), 6-9.

Abstract

We describe a chemical tag for duplex proteome quantification using neutron encoding (NeuCode). The method utilizes the straightforward, efficient, and inexpensive carbamylation reaction. We demonstrate the utility of NeuCode carbamylation by accurately measuring quantitative ratios from tagged yeast lysates mixed in known ratios and by applying this method to quantify differential protein expression in mice fed a either control or high-fat diet.



Introduction.

Recently we described a fresh approach to MS1-based proteome quantification by metabolic labeling that increases plexing capacity by compressing the “light” and “heavy” partner spacing by ~ 103, i.e., from several Da to several mDa.[22] To achieve mDa spacing we grew cells on two isotopologues of lysine – each having eight additional neutrons, one including ^{13}C and ^{15}N and the other with 2H. Here “light” and “heavy” peptide partners were spaced by 36 mDa and are distinguished from one another with MS1 mass resolving powers in excess of ~ 100,000. Such resolution is readily accessed with modern Fourier Transform mass analyzers such as Ion Cyclotron Resonance (FT-ICR) and Orbitraps. Because the method exploits the subtle differences in neutron binding energy of the various stable isotopes we called this method neutron-encoded quantification (NeuCode). Note the phenomenon of mass defect has been previously exploited in proteomics. Prior uses include peptide labeling with amine- or cysteine-reactive tags containing heavier isotopes, such as ^{79}Br, to impart large and unique mass defects and move peptide ions into regularly unoccupied spectral space.[21] Orlando and colleagues used mass defect by permethylating glycans with labels incorporating either ^{1}H or ^{2}H. Because each glycan accepted several labels mass differences in excess of 50 mDa could be induced and distinguished with MS resolving powers of 25,000.[36] Building on these ideas, NeuCode uses multiple isotopes (C, H, N, and O) and typically engineers mass differences ranging from 6 to 36 mDa into peptides and requires resolving powers in excess of 100,000.

Here we expand the NeuCode method by the design of a simple and readily accessible chemical tag. Certain samples, including biological tissues and fluids, are challenging to label metabolically, hence, chemical approaches to introduce stable isotopes are key to many applications. Further, if well-chosen, reagents for chemical labeling can be quite inexpensive. We selected urea (CH_4N_2O) as it can carbamylate free amines with excellent efficiency under simple conditions (i.e., urea + heat). For each site of carbamylation on a peptide, typically two

for proteins digested with LysC, a 6.3 mDa mass difference is encoded. Here, we evaluate the quantitative accuracy and precision of these duplex NeuCode chemical tags using complex mixtures of proteins from yeast and mouse tissues.

Methods

CAST/EiJ mice (Jackson Laboratory, Bar Harbor, Maine) were fed a control diet or a high fat/high sucrose diet for 5 months. Proteomic samples from either mouse liver or yeast cultures were performed as previously described.[22,37,38] Isotopically labeled urea ($^{15}N_2$ or $^{13}C_1$) was purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Peptides were carbamylated by incubating at pH 8 with a large excess of aqueous 8 M urea at 80˚C for 4 hours and remaining urea was removed by desalting (Online Resource 1).[39] Fractionation of the murine peptides was performed by offline high pH reverse-phase separation.[37] Peptide mixtures were separated using self-prepared capillary LC columns and analyzed upon elution using an Orbitrap Elite mass spectrometer with conditions described elsewhere (Thermo Scientific, San Jose, CA).[22] **Figure 2.1** presents an overview of the scan methods. Resulting data was searched against a target-decoy database (1% FDR) using the Open Mass Spectrometry Search Algorithm (OMSSA).[40-42] Proteins were quantified by extracting the sum of all contributing peptide intensities using a custom, in-house algorithm (manuscript in preparation). Abundance ratios were normalized to account for differences in total material.

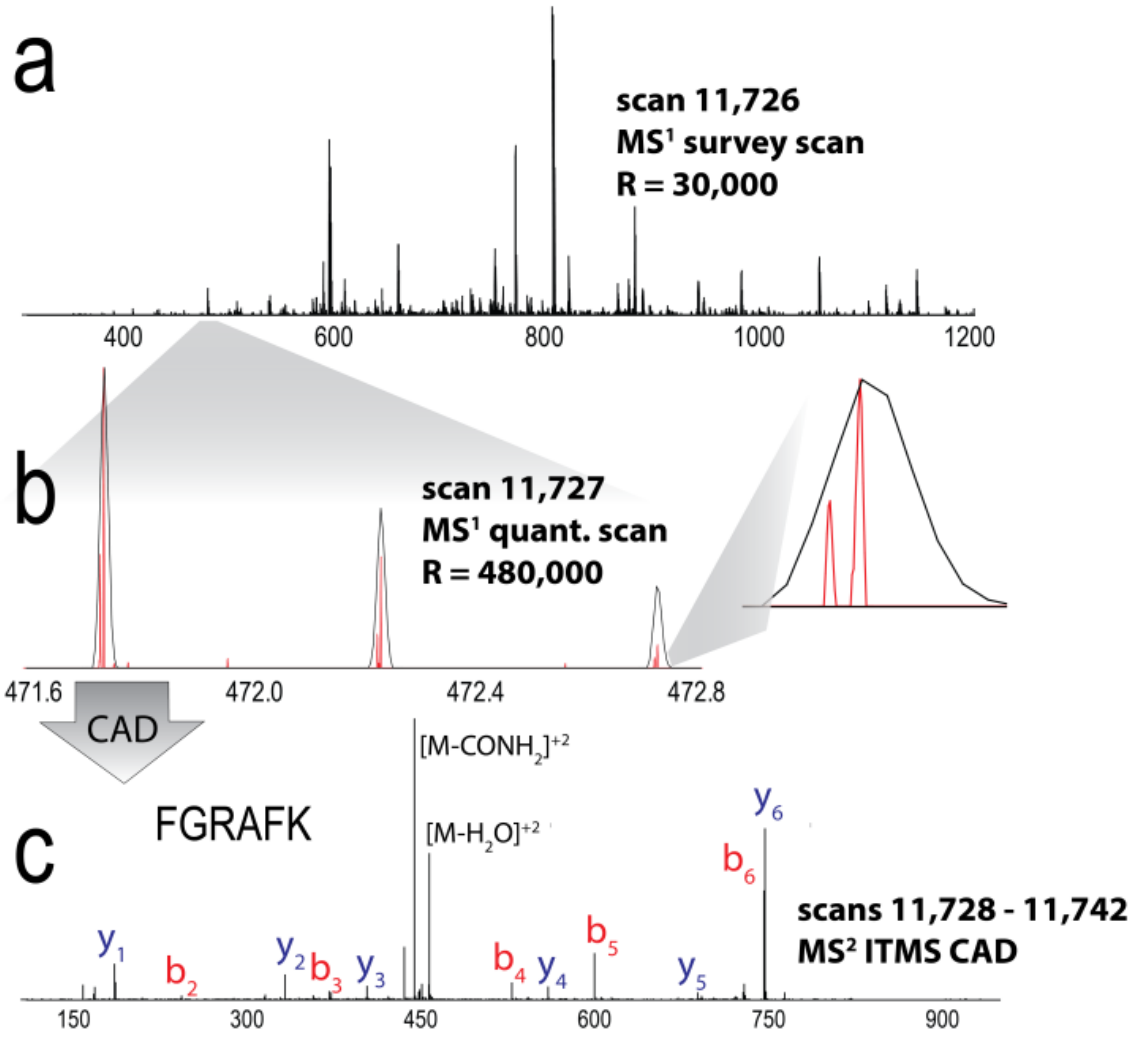


Figure 2.1. Data acquisition method for analysis of carbamyl NeuCode tagged peptides. **A** Orbitrap MS^1 survey scan (R = 30,000). **B** Orbitrap quantitation scan (MS^1, R = 480,000) reveals the quantitative information. **C** Ion trap CAD MS/MS scan of carbamyl NeuCode tagged peptide having the sequence FGRAFK

Results

Carbamylation with isotopically light and ^{13}C and ^{15}N enriched urea, which incorporates a 2 Da mass difference, has been previously described for MS^1-based protein quantification.[39] Building on this research, we selected the carbamylation reaction for NeuCode labeling for three reasons. First, carbamylation of amines with urea is straightforward, simple, and well-established. Second, carbamylation only modestly increases overall peptide mass (44 Da). Third, isotopic versions of urea are readily available and relatively inexpensive.

Theoretical Background.

We assembled a library of 36,783 experimentally detected peptide ions – using their observed charge (*z*) and mass-to-charge (*m/z*) ratios for *in silico* calculations. The calculations counted the number of free amines and installed a corresponding number of carbamyl moieties in either NeuCode heavy (^{13}C) or light (^{15}N) flavors. We calculated whether each of the NeuCode partner peaks could be separated (full-width at 10% of maximum, FWTM) at resolving powers up to 10^6 (**Figure 2.2 A**). This relatively conservative definition only considers NeuCode peak areas overlapping by less than 3.2% as quantifiable. Given that commercial Orbitrap hybrid MS systems can achieve routine resolution of 480,000 (@ m/z 400), we conclude that the majority of peptides labeled by NeuCode carbamylation (71%) are suitable for quantification. Further improvements to Orbitrap analyzer technology, or analysis on a higher resolving power FT-ICR system, will increase the quantifiable percentage. Peptides suitable for quantitation had an average *m/z* of 698 and an average charge state of 2.41, whereas peptides unsuitable for quantitation at the employed resolving power had an average *m/z* of 930 and an average charge state of 3.02. Nearly all the carbamyl NeuCode doublets are separable at the highest reported Orbitrap resolution of 10^6.[43] Further, FT-ICR devices achieve the highest resolution and could likewise resolve virtually all peptide precursors carrying such labels.

Reaction and Label.

Following carbaminomethylation of cysteine residues, we digested a complex mixture of yeast proteins using LysC and subjected the resulting peptides to a carbamylation reaction with either ^{13}C or ^{15}N labeled urea.[39] We achieved labeling efficiencies in excess of 95%. Only deprotonated amines can be carbamylated. Consequently, R residues will not undergo this reaction at pH 8 (pKa = 12.5). Little evidence for R or H carbamylation was evident - 0.9% of R residues (37/3931) and 1.1% of H residues (24/2092) were carbamylated (1% FDR). Note peptides resulting from LysC digestion accept two carbamyl groups and, thus, incur a mass difference of 12.6 mDa. Further, LysC digestion provides comparable coverage of the yeast proteome as compared to trypsin.[44] The carbamyl NeuCode tagged peptides were loaded onto a nanoflow capillary reversed-phase LC column and were ionized by an integrated electrospray emitter. The resulting protonated peptide ions were mass analyzed using an Orbitrap hybrid MS. The system conducted an initial FT-MS survey scan (30,000 resolving power) to direct data dependent MS/MS scans followed by a high resolution MS^1 scan (R = 480,000 @ *m/z* 400) for quantitation (**Figure 2.1**). Note this resolution is available on all Orbitrap Elite MS systems by installation of the manufacturer's developer's kit software. MS/MS scans were acquired using the linear ion trap (**Figure 2.1 C**). As evidenced in **Figure 2.1**, the quantitative information is concealed in both the initial MS^1 survey scan (R = 30,000 @ *m/z* 400) and in the ion trap MS/MS scans. The high resolution quantification scan, however, separates the NeuCode pair, revealing the quantitative data. The high resolution MS^1 scan is conducted concurrently with the trap MS/MS events to optimize duty cycle.

Effects of high resolution scan cycle.

The NeuCode approach combines advantages of other quantitative methods. Similar to metabolic labeling approaches, each peptide's natural isotope pattern provides multiple isotopic peaks for garnering quantitative measurements. Additionally, as in isobaric tagging strategies,

the signal across quantitation channels is combined during survey MS^1 scans so that redundant MS/MS scans are not acquired by selection of identical peptide sequences across multiple isotopic clusters. This method thus affords a larger signal and lower spectral complexity for deep sampling of the proteom.[45] Finally, the MS/MS-independence of quantitation of the NeuCode method obviates the need for high precursor ion purity.

The NeuCode scan cycle requires the collection of a high resolution MS^1 scan. Today the acquisition of a 480K scan requires collection of a 1.7 second transient. Doubtless the duration required to achieve this resolving power will be reduced as instrumentation continues to evolve; that said, by parallel collection of MS/MS scans, only a very subtle decrease in overall duty cycle is imposed. To document this we injected a digested yeast protein lysate and analyzed it using both the NeuCode scan sequence and a traditional data-dependent method. Despite the addition of a high-resolution FTMS quantitation scan, only 4% fewer peptides were identified from the data resulting from the NeuCode analysis, i.e., 9,276 instead of 9,685 (1% FDR).

Quantitative accuracy. Next we investigated quantitative accuracy – a critical figure of merit of any quantitation method. We combined aliquots of labeled yeast LysC peptides in 1:1 and 5:1 ratios before LC-MS/MS analysis. Each sample was interrogated in triplicate. The median measured ratios for identified peptides were 1.03:1, 1.03:1, and 1.03:1 and 5.07:1, 5.06:1, and 5.03:1. Standard deviations were between 0.21 - 0.23 and 0.52 - 0.61, respectively, on a log2-fold change scale (**Figure 2.2 D**).

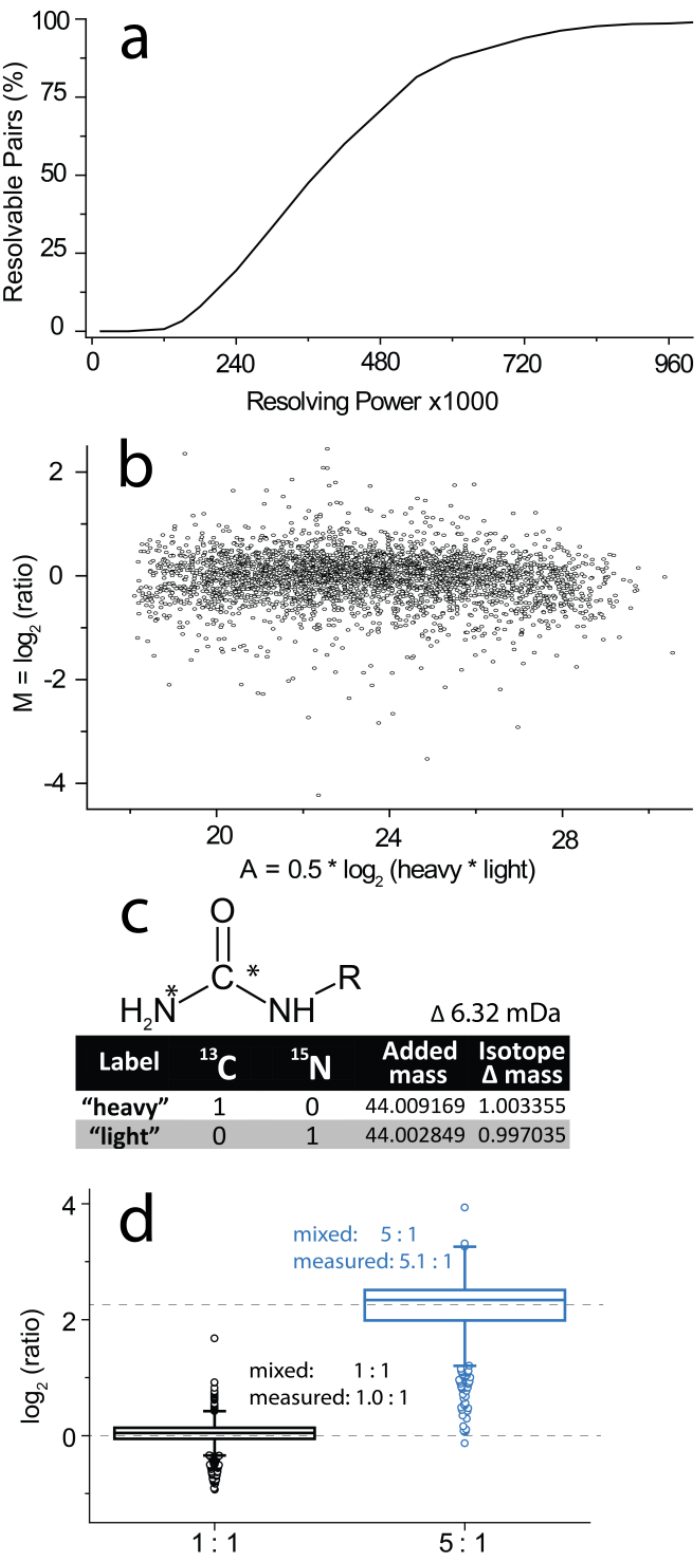


Figure 2.2. Overview of the carbamyl NeuCode strategy. **A** Theoretical calculation of the percentage of carbamyl NeuCode tagged peptides that are resolvable, thus, quantifiable, at resolutions up to 10^6 (full width at 10% max). **B** Fold change (M) for every protein in the murine high-fat diet model versus total intensity (A) of all peptides mapping to a protein. **C** The urea label, isotope substitutions, and mass changes employed to achieve carbamyl NeuCode reagents. **D** Box plots showing the measured and true 1:1 and 1:5 peptide ratios. The box represents the median, 25th, and 75th percentile while the whiskers show 1.5 x interquartile range. The circles show outliers

High fat diet induced proteome changes.

To further validate the NeuCode method we utilized a mouse model - a typical tissue sample type that is challenging to label using metabolic approaches. We measured the effects of a high fat diet on the proteome of mice from the CAST/EiJ strain, which are highly resistant to diet-induced obesity and diabetes. Liver tissue was homogenized and the cells lysed. Whole protein extracts were digested with LysC, labeled with carbamyl NeuCode tags, fractionated, and measured by nHPLC-MS/MS. In total, we identified 14,718 murine peptides. Of all theoretically separable NeuCode partners, 87% were experimentally quantified. The slight reduction from theory is mostly due to a subset of low signal-to-noise precursors; this is likewise observed in traditional SILAC experiments. Note that closely spaced *m/z* peaks, such a NeuCode partners, can sometimes be detected as a single entity due to the well-known phenomenon of peak coalescence. We did notice occasional coalescence of the most intense peaks in the high resolution MS^1 scans. In those cases our data analysis software uses lower abundance isotopes to extract quantitative information (as described elsewhere).[22]

In accordance with previously published research, proteins associated with peroxisomes were considerably up-regulated in the mice that were fed a high-fat diet.[46] Further, expression of mitochondrial proteins was broadly down-regulated.[47] These findings were established by comparing the population of protein groups up-regulated by at least 50% (184) and protein groups down-regulated by at least 50% (337) to the total population (n = 2,879) of quantified proteins.[48] Finally, we conclude these detected differences were not due to increased variability of lower signal datapoints by comparing the measured fold-change to signal intensity (**Figure 2.2 B**).

Conclusions

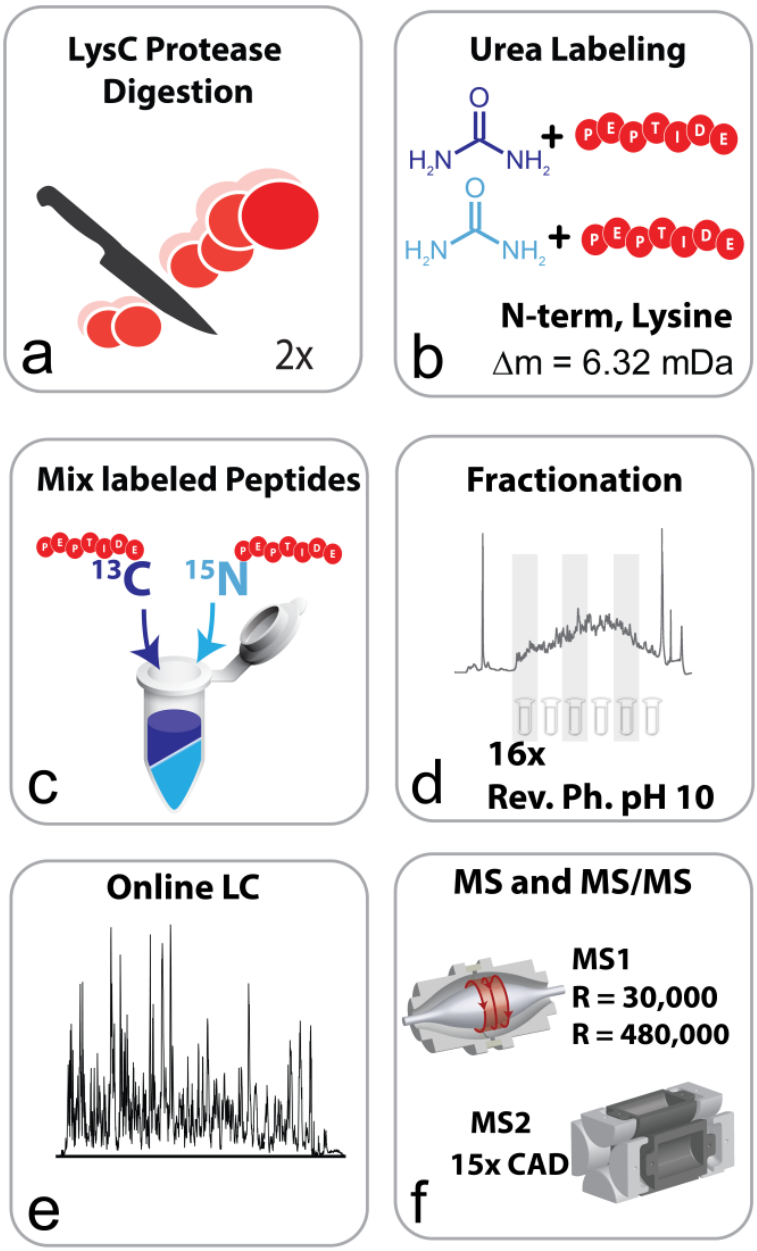
We have demonstrated the feasibility of NeuCode carbamylation via urea isotopologues for relative proteome quantification. The labeling reaction is reproducible, highly efficient, and relatively simple, requiring only a urea solution and a heated reaction vessel. Experimentally measured relative abundances of peptides from a combined yeast digest mixed in known ratios were very close to the true ratios. The major requirement to implement NeuCode urea labeling is a mass spectrometer capable of a resolving power of ~ 500,000 (@ 400 *m/z*). Current and recent models of Orbitrap MS systems already achieve this performance in commercial platforms. Using ^{15}N and ^{13}C-urea-based labeling affords a robust, accurate, and tissue-compatible implementation of duplex NeuCode quantitation.

Acknowledgment

This work was supported by the National Institutes of Health grants R01 DK066369, DK058037 and DK091207 (A.D.A.), and GM080148 (J.J.C.). A.E.M. gratefully acknowledges support from a National Institutes of Health-funded Genomic Sciences Training Program (5T32HG002760).

Supporting Information Available

A Supplemental Figure describing sample preparation appears on the next page. An Excel spreadsheet showing theoretical calculations, and an Excel spreadsheet showing measured protein fold-changes in the mouse model were available online with the published paper but could not be included in this thesis. The spreadsheets are accessible through PubMed Central® at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3880647/.



Supplemental Figure 2.2.1. Sample preparation overview for carbamyl NeuCode method. Enzymatic digestion **A** of proteins is followed by labeling with heavy urea **B**. Labeled peptides are mixed **C** and offline fractionation was undertaken **D.** Finally, peptides were interrogated by LC-MS/MS **(E+F)**

Chapter 3: NeuCode enabled Quantitation of Organic Acids by LC-MS and GC-MS

A.U. designed experiments, performed experiments, analyzed data, and wrote the paper.

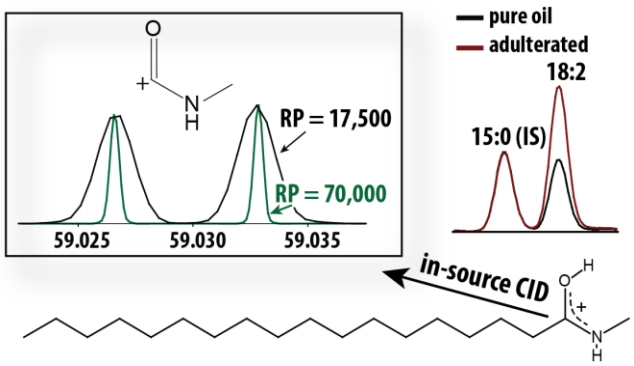
Part of this chapter has been published:

Ulbrich, A., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2014). Organic acid quantitation by NeuCode Methylamidation. *Analytical Chemistry*, *86*(9), 4402-4408

3.1: Organic acid quantitation by NeuCode methylamidation

Abstract

We have developed a multiplexed quantitative analysis method for carboxylic acids by liquid chromatography high resolution mass spectrometry. The method employs neutron encoded (NeuCode) methylamine labels (^{13}C or ^{15}N enriched) that are affixed to carboxylic acid functional groups to enable duplex quantitation via mass defect measurement. This work presents the first application of NeuCode quantitation to small molecules. We have applied this technique to detect adulteration of olive oil by quantitative analysis of fatty acid methyl amide derivatives, and the quantitative accuracy of the NeuCode analysis was validated by GC-MS. Currently the method enables duplex quantitation and is expandable to at least 6-plex analysis.



Introduction

Gas chromatography coupled with mass spectrometry (GC-MS) is a standard technology for food analysis, clinical diagnostics, and drug development and allows for detection and quantification of various classes of compounds. Since the 1980s, widespread use of liquid chromatography mass spectrometry (LC-MS), enabled by electrospray ionization (ESI), has increased the throughput and breadth of molecules that can be mass analyzed.[45,49-53] For example, food products are routinely screened by GC- and LC-MS to discover and quantify contaminants and ensure authenticity.[54-58] Detecting adulteration of food is important for public health and requires accurate quantitative measurements, with LC-MS currently being the primary method of analysis.[59] The standard workflow involves separating molecules derived from food products by LC and mass analyzing them one sample at a time. Comparative measurements can be made when the intensities of matched peaks are compared between multiple LC-MS experiments. Strategies for increasing the throughput of LC-MS for protein samples have involved multiplexed quantitative analysis by monitoring several samples simultaneously.[60-62] Proteomic LC-MS analysis has led the way in increasing quantitative throughput by using stable heavy isotopes as a means to differentiate samples. Incorporating heavy isotopes, e.g., ^{13}C, ^{15}N, or ^{2}H, into peptides through metabolic or chemical means, creates mass differences that are used to quantify relative abundances among samples. This mass difference is often designed to be several Da (e.g., 4 or 8) in size to minimize isotopic overlap.[45,63-65] Small molecule LC-MS analysis has lagged behind proteomics in widespread adoption of multiplex quantitation techniques. This is in part due to the shorter analysis times (< 30 minutes) compared to proteomics, but also to the cost of quantitative labels and their added mass. Still, some differential labeling approaches do exist for small molecule LC-MS analysis. Recent techniques use cholamine derivatives of carboxylic acids ($^{2}H_0$ or $^{2}H_9$), N-hydroxynicotinic

acid derivatives of Vitamin E related compounds ($^{2}H_0$ or $^{2}H_3$), and dansylation ($^{13}C_0$ or $^{13}C_2$) of alcohol and amine metabolites successfully for duplex quantitation.[66-68]

Recently, we developed a quantitative proteomic method that uses mDa mass differences that arise from varied nuclear binding energies. This approach, called neutron encoded quantitation (NeuCode), exploits a small difference in added mass (3 to 36 mDa) between ^{13}C, ^{15}N, or ^{2}H atoms.[16,22,69-71] The resulting mass difference is much smaller than in traditional isotope enriched samples and, consequently, the size and mass of the quantitative label can be small while avoiding overlapping isotopic distributions.[16] Here we have expanded NeuCode quantitation to small molecule analysis by derivatizing carboxylic acids with two forms of heavy methylamine ($^{13}CH_5N$ and $CH_5{}^{15}N$) that have a mass difference of 6.32 mDa. To resolve this narrow spacing and quantify samples, high resolution mass analyzers are needed (*e.g.,* Orbitrap, FT-ICR, time-of-flight (TOF)). However, due to smaller molecular weights and lower charge states of the analytes, resolution requirements are lower than in NeuCode proteome analysis.[21] Here we describe the quantitation of carboxylic acids, particularly fatty acids in pure and adulterated olive oil, using methylamine NeuCode labels via LC coupled to a quadrupole-Orbitrap mass spectrometer. Quantitation was confirmed by a well-established GC-MS method.[72] We concluded that methylamine is an ideal NeuCode label since it (1) imparts a small delta mass (+15 Da), (2) increases response in positive ESI, and (3) heavy isotopologues of methylamine are commercially available. We show that either the $[M+H]^+$ ion or a number of product ions arising from collisional activation can be used for quantitation.

Experimental

Materials and Reagents

Benzoic acid, thionyl chloride, methylamine hydrochloride, and fatty acid standards were obtained from Sigma Aldrich (St. Louis, MO) and used without further purification. Isotopically

labeled methylamine hydrochloride containing either > 98% ^{15}N or > 99% ^{13}C was obtained from Cambridge Isotope Laboratories (Tewksbury, MA). A standard mixture of free fatty acids was obtained from Larodan Fine Chemicals (Malmö, Sweden).

Hydrolysis and Labeling Procedure

Fifty µL of olive oil (HyVee Select, Classic Olive Oil, Des Moines, IA) were hydrolyzed in 1000 µL 0.5 N HCl in 9:1 acetonitrile : water at 100˚C for 30 min.[73] The adulterated sample contained 30% Crisco Vegetable (soybean) Oil (The J.M. Smucker Company, Orrville, OH). Three mg of pentadecanoic acid were added to the solution as an internal standard. Oxidation was minimized by sparging acetonitrile and water with nitrogen and freshly preparing the HCl solution before each use. The reactions were carried out in 8 mL screw-cap glass centrifuge tubes (VWR, Radnor, PA) and the reaction mixtures were blanketed with nitrogen before incubation in a hot water bath. After hydrolysis, 1000 µL chloroform and 1000 µL water were added, the reaction mixture was shaken, and the organic phase was kept. The chloroform extract was dried in vacuo and the residue redissolved in 1 mL methylene chloride. Then 50 µL thionyl chloride and 2 µL catalytic pyridine were added and the lightly capped tubes, to allow escape of gaseous reaction products, were shaken for 5 min.[74] Excess thionyl chloride and methylene chloride were evaporated in vacuo. To the obtained acyl chlorides were added 250 µL methylene chloride and then 100 µL of 20% aqueous NaOH containing 15 mg methylamine hydrochloride. The mixture was vigorously agitated for 1 min, 2 mL water were added, and then it was extracted with 2 mL methylene chloride. The extract was washed with 2 mL water. When cloudiness persisted in the methylene chloride layer 500 µL methanol were added to clarify the organic layer, which will remain at the bottom.

For the derivatization and analysis of benzoic acid and free fatty acids the hydrolysis step was omitted and acid aliquots (10-20 mg) were activated with thionyl chloride and then reacted with three- to four-fold molar excess of methylamine hydrochloride as described above.

LC-MS Analysis

LC-MS analysis was performed using an Ascentis Express reverse phase C18 column (150 mm x 2.1 mm x 2.7 µm particle size, Supelco, Bellefonte, PA) and an Accela LC Pump (300 µL/min flow rate, Thermo Scientific, San Jose, CA). Mobile phase A consisted of 10 mM aqueous formic acid buffered to pH 3 with ammonia and mobile phase B consisted of 10 mM methanolic formic acid containing 75 µL/L 30% aqueous ammonia. Initially, mobile phase B was held at 35% for 1 min and then linearly increased to 75% over 4 min and then to 100% over 13 min. At 23 min the mobile phase composition was returned to initial conditions (35%) over 0.1 min and held for 4.9 min for column re-equilibration. Note the method only requires methanol and water as solvents avoiding the use of more hazardous organic solvents. Oil and organic acid methyl amide samples were diluted in methanol and 5 µL were injected by an HTC PAL autosampler (Thermo Scientific, San Jose, CA). The LC system was coupled to a Q Exactive mass spectrometer (Build 2.3 SP2) by a HESI II heated ESI source (Thermo Scientific, San Jose, CA). The inlet capillary was kept at 300˚C, sheath gas was set to 35 units, and auxiliary gas to 10 units. The MS was operated in positive ion mode with the full MS mass range spanning 185 to 335 Th at resolving power (RP) 140,000 or 70,000. A full MS AGC target of 5e4 was used. Quantitation was performed by analyzing $[M+H]^+$ ions or ions produced from in-source fragmentation (80 eV) using 17,500 RP and a mass range from 57 to 63 Th.

Quantification of LC-MS results

Absolute quantification proceeds using integrated peak areas and linear (equally weighted) calibration curves (Xcalibur software suite, Thermo Scientific, San Jose, CA). Calibration curves of fatty acid methyl amides were constructed from eight injections of a 1:2 serial dilution of a fatty acid amide mix. Concentrations ranged from 0.1 ng / mL for the highest dilution of the least abundant fatty acid amide (18:3) to 195 ng / mL for the high standard of the most concentrated fatty acid (18:1). Relative quantitation uses ratios of integrated peak areas of extracted ion

chromatograms of NeuCode analyte peaks. The difference in isotopic purity of the used methylamine reagents was determined by LCMS and a correction factor of 1.0061 was introduced due to the lower isotopic purity of ^{15}N-enriched compared to ^{13}C-enriched methylamine (98.6% vs. 99.3%). Chromatograms underwent 7-point box-car smoothing before calculation of quantitative ratios to facilitate automated peak integration but no other scan averaging was used. Confirmation of quantitative results was performed on a GC-TOF-MS system (Pegasus 4D GCxGC-TOFMS, Leco, St. Joseph, MI) using ChromaTOF software (Leco, St. Joseph, MI) by analysis of the fatty acid methyl esters following established methods.[72]

Simulations

Resolvability simulations at different intensity ratios, *m/z* values, and resolving powers were developed and visualized using MathCad 14 software (PTC, Needham, MA). The workbooks are available as an online supplement to this article.

Safety Considerations

Thionyl chloride is corrosive, incompatible with water and alcohol solvents, and its decomposition products are toxic. This chemical should only be used in a hood and protective clothing and goggles should be worn when handling it. The 20% NaOH solution is corrosive and should be handled with care while wearing gloves and eye protection. Pyridine is toxic and malodorous and should be used in a hood while wearing gloves and goggles.

Results and Discussion

NeuCode Quantitation

NeuCode quantitation relies on resolving two (or more) *m/z* features that are closely spaced. The main challenge, therefore, is the increased resolving power demand to adequately separate these small differences. Using methylamine NeuCode labels, only a 6.32 mDa mass difference

occurs between the ^{13}C and ^{15}N isotopologues. To determine if a quadrupole-Orbitrap MS (Q Exactive), which does not contain a high-field Orbitrap, has the necessary resolving capabilities, we calculated which *m/z* values could be distinguished as a function of resolving power (RP). Approximating FTMS *m/z* peak shapes as Gaussian, we declare that two equally intense peaks are resolvable when the difference between their measured *m/z* values is greater or equal to their full width at ten percent maximum height (FWTM). **Equation 3.1.1** calculates the theoretical *m/z* width of a peak at a given *m/z* and analyzed at a RP (defined at 200 *m/z* for the Q Exactive)

$$\Delta \frac{m}{z}_{theo} = \frac{\sqrt{2\ln f}}{\sqrt{2\ln 2}} \frac{(m/z)}{RP\sqrt{200/(m/z)}} \qquad \text{eq. } 3.1.1$$

The *f* -term represents the fraction of maximum height (e.g., 10) and the denominator is the effective resolving power at the specified *m/z* value for an Orbitrap mass analyzer. To determine which *m/z* values are resolvable on the quadrupole-Orbitrap given a 6.32 mTh NeuCode spacing at FWTM, we solved **equation 3.1.1** for (m/z) where:

$$\Delta \frac{m}{z}_{theo} \geq 6.32\ mTh \text{ and } f = 10$$

and arrived at

$$(m/z)_{resolvable} \leq 0.134\ Th\ x\ RP^{2/3}$$

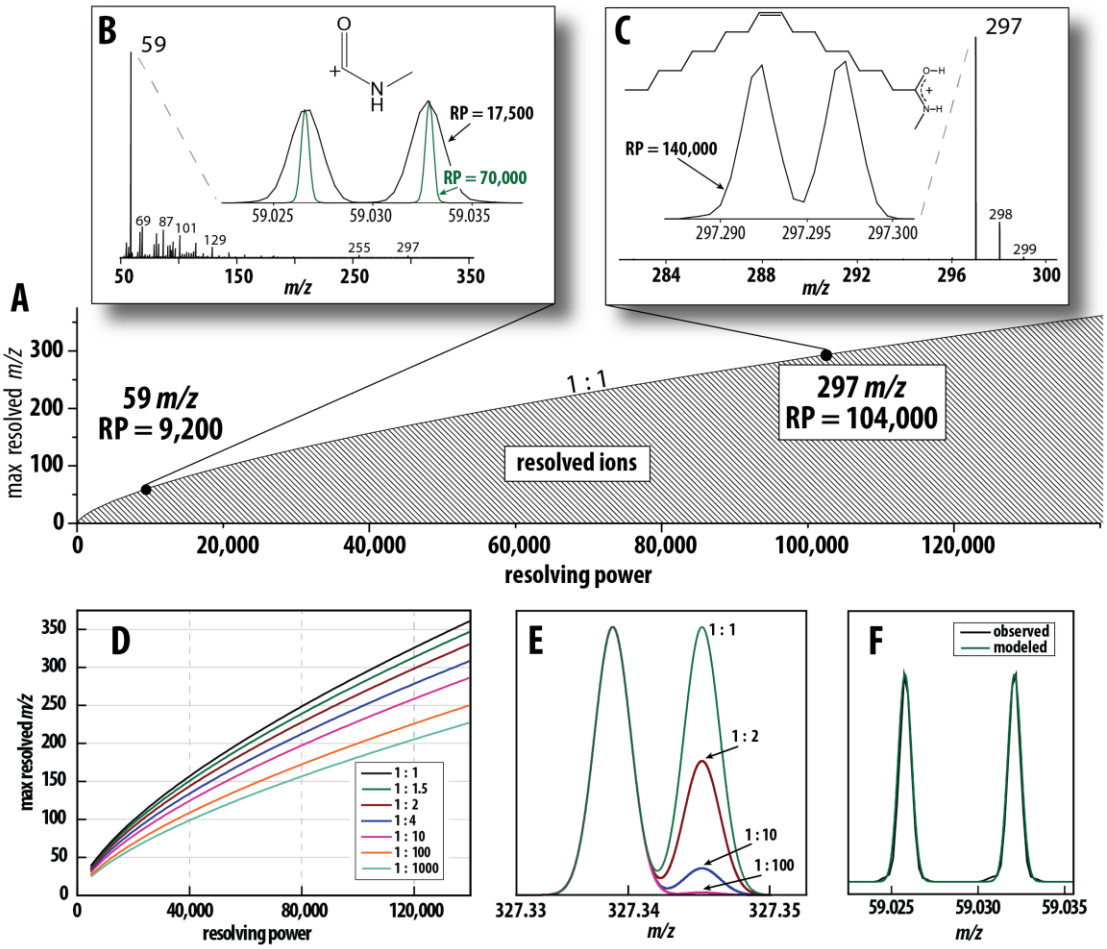


Figure 3.1.1. Required resolving powers for NeuCode quantitation. **A** Resolvable ions as a function of instrumental resolution. All *m/z* values below the line can be resolved. **B** Mass spectrum of oleic acid methyl amide (source fragmentation). The fragment ion at 59 *m/z* is shown in magnification. The proposed structure of the fragment is shown above the doublet of peaks. **C** Full scan mass spectrum of oleic acid. The quantitative pair shown requires the much higher RP 140,000 to separate the partners. **D** The effect of relative concentrations of analytes on maximum resolvable *m/z* at various quantitative ratios. The maximum quantifiable *m/z* decreases with an increases of the ratio. **E** Simulation of the NeuCode peaks of differentially labeled arachidic acid at different ratios. **F** An overlay of observed and simulated NeuCode peaks. The empirically observed peaks have been corrected for systematic mass error only.

Equation 1 is plotted in **Figure 3.1.1 A** and shows which *m/z* values of NeuCode pairs, at a 1:1 ratio, can be separated at a given resolving power. For example, an ion at 59 *m/z* only requires at least 9,200 RP, whereas an ion at 297 *m/z* requires greater than 104,000 RP (**Figure 3.1.1 B, C**). The highest resolvable *m/z* value at the maximum quadrupole-Orbitrap RP of 140,000 is 361 *m/z*. This enables fatty acid methyl amides up to C22:0 (355 *m/z*) to be quantified by their $[M+H]^+$ ions at the maximum RP. Fortunately, most carboxylic acid metabolites have molecular weights well below 361 Da and can be quantified at lower resolving powers.

Quantitation of analytes at significantly different concentrations requires resolving peaks of different intensities (**Figure 3.1.1 D**). The resolving power demands increase and the highest quantifiable *m/z* decreases with growing ratios (**Figure 3.1.1 E**). For example, resolving a pair of peaks at a 1:10 intensity ratio is possible only up to 286 *m/z* at 140,000 RP compared to a maximum of 361 *m/z* when the ratio is 1:1. Since the abovementioned values stem from our Gaussian model, we wanted to ensure that our model accurately reflects actual data. We overlaid simulated and observed NeuCode pairs and found sufficient agreement between our model and empirical data (**Figure 3.1.1 F**).

Quantifying NeuCode samples has typically been performed on intact precursor ions detected in a full MS scan, but quantitation can also be derived from product ions, provided they contain the encoded neutrons.[22,69] Fatty methyl amides produce neutron encoded product ions when fragmented by in-source or beam-type collisional activation (HCD). A prominent fragment of all methyl amides is found at 59 *m/z*. Its structure (**Figure 3.1.1 B**) shows that both neutron-bearing sites are present and, thus, it contains the quantitative information. Additionally, quantifying product ions of low *m/z* eases the resolving power requirements to sufficiently separate them; resolving powers as low as 10,000 are capable of completely separating these NeuCode peaks at 59 *m/z* (**Figure 3.1.1 A**). Note this resolving power is obtainable on non-FTMS instruments, e.g., TOFs, and enables NeuCode quantitation for a wider range of mass

spectrometers. For FTMS instruments, resolving power is inversely proportional to either the square root of the *m/z* (Orbitraps) or the *m/z* (FT-ICR), and since resolving power is also proportional to acquisition time, quantifying lower *m/z* small molecules by NeuCode is also faster than quantifying peptides. This leads to increased scan rates, allowing collection of more data points across elution profiles for better quantitation. To demonstrate the applicability of using lower resolutions, oleic acid was fragmented by in-source fragmentation (80 eV) and the fragments were mass analyzed at 17,500 and 70,000 RP (**Figure 3.1.1 B).** A zoom-in on the characteristic even-electron ion at 59 *m/z*, which bears the encoded neutron and arises from α-cleavage, shows that both resolving powers are capable of baseline resolution of the NeuCode pairs and that the quantitative results are identical.[36] The only difference is that the 70,000 RP scan took four times as long to acquire. In contrast, analysis of the $[M+H]^+$ ion at 297 *m/z* requires a RP of 140,000 to fully separate and quantify oleic acid, taking eight times as long (**Figure 3.1.1 C)**. Additionally, since all fatty acid methyl amides produce the ion at 59 *m/z*, any length fatty acid can be quantified, given baseline chromatographic separation. This low mass fragment is also observed when other methyl amides, such as methyl benzamide, are fragmented and can serve as a universal quantitative ion for analysis of organic acids via their methyl amides. We conclude that NeuCode labels with a Δ mass of 6.3 mDa on carboxylic acids can be quantified with instrumental resolving powers as low as 10,000.

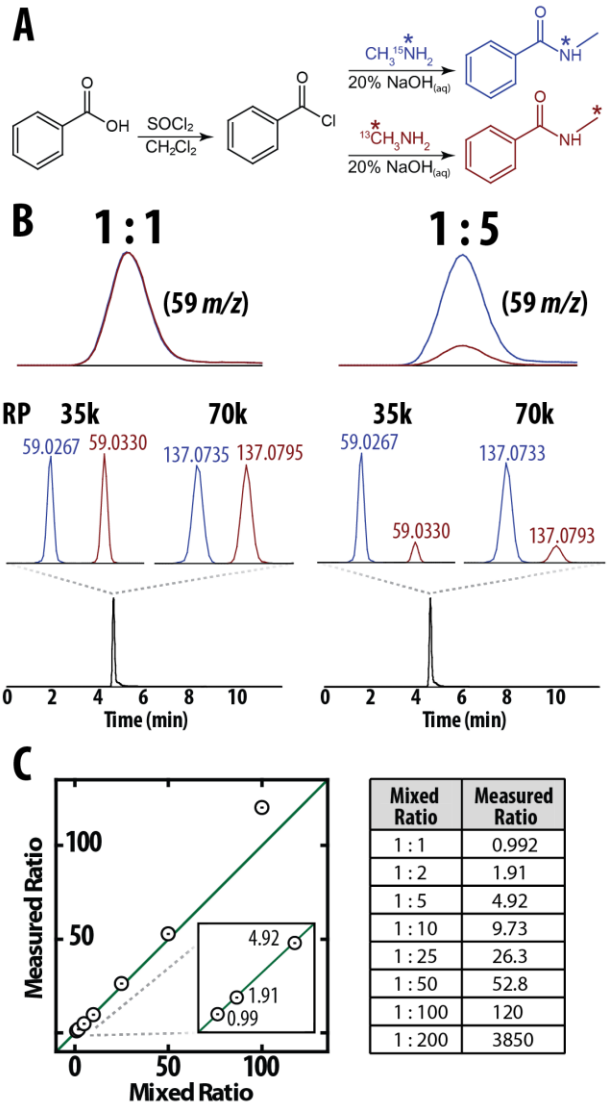
Reaction and Label

We selected methylamine as the NeuCode reagent to produce fatty amides for three reasons: it is (1) very small, relying on a single ^{13}C and ^{15}N, (2) readily available as isotopologues, and (3) increases ionization efficiency in positive ESI. When used as a duplex (^{13}C and ^{15}N) NeuCode label, methylamine does not contain deuterium and therefore no detectable chromatographic shift occurs. Additionally, linking methylamines to fatty acids is a straightforward reaction and proven reagents are readily available.[37] We chose the amide bond formation route via the acyl

chloride intermediate (**Figure 3.1.2 A**) over the use of coupling reagents for two reasons: (1) acyl chloride formation by thionyl chloride is robust and (2) reaction side products are either volatile (SO_2, HCl) or benign (H_2O).[38] Note that activation of the carboxylic acid is necessary since direct reaction of the carboxylic acid with methylamine will merely give the methylammonium carboxylate salt instead of the methyl amide.

Labeling and analysis of benzoic acid and a mixture of fatty acids

We first tested the labeling procedure on a solution of benzoic acid, split in two and labeled with either the heavy ^{13}C (32.04555 Da) or light ^{15}N (32.03923 Da) methylamine, and then recombined at 1:1 or 5:1. The reaction gave an 84% final yield, and LC-MS/MS analysis (RP = 70,000) of the reaction products showed one major peak at the expected methylbenzamide *m/z* value of 137.1 (**Figure 3.1.2 B**). Three replicate derivatizations were carried out and were highly reproducible with a relative standard deviation of the reaction yield of 2.1%. Note, however that the amount of material in each quantitative channel is normalized by inclusion of an internal standard before derivatization (pentadecanoic acid) for duplex experiments to correct for slight differences in yield. Extracted ion chromatograms (XICs) of the light and heavy $[M+H]^+$ ions show measured ratios of 1.0:1 and 5.8:1 with no detectable chromatographic shift. When fragmented, methylbenzamide produced the characteristic ion at 59 *m/z*, which gives similar (1.0:1 and 5.6:1) quantitative results compared to the intact precursor (**Figure 3.1.2 B**). This validates that using the peak at 59 *m/z* can serve as the quantitative proxy for the precursor without degraded quantitative accuracy. In another set of experiments, we combined solutions of differentially labeled methylbenzamide at a variety of ratios before injection to compare the measurements to the known mixing ratios. Relative concentrations up to 50:1 could be accurately quantified while quantitation ability was lost entirely above a 100:1 ratio as the lower channel disappeared into the noise band (**Figure 3.1.2 C**).



Mixed Ratio	Measured Ratio
1 : 1	0.992
1 : 2	1.91
1 : 5	4.92
1 : 10	9.73
1 : 25	26.3
1 : 50	52.8
1 : 100	120
1 : 200	3850

Figure 3.1.2. Overview of the duplex NeuCode quantitation method for carboxylic acids via their methyl amides. **A** Reaction scheme showing the preparation of the ^{15}N and ^{13}C labeled methyl amides from benzoic acid. **B** HPLC-MS elution profiles of 1:1 (measured at 1:1.02) and 5:1 (measured at 1:5.55) mixtures of light and heavy methylbenzamide, mass peaks of the NeuCode pairs at the $[M+H]^+$ ion, and the NeuCode fragment at 59 *m/z*. **C** The measured ratio as a function of the known ratio of isotopically labeled methylbenzamides. The green line represents complete agreement between known and measured ratios. Quantitative results approximate true mixing ratios closely up to a ratio of 50:1.

Ratios were calculated by dividing the areas under the curve (AUC) of the elution profiles of the NeuCode ions, which consisted of about 30 points across each elution profile (4 σ width, ~20 s) at 140,000 RP or upwards of 100 points at 35,000 RP. We observed that peaks having 7-second chromatographic peak widths, typical in UHPLC separations, still provided 50 points over the elution profile at 35,000 RP and 85 points at 17,500 RP. NeuCode quantitation can thus keep pace with fast separations, especially when lower RPs are used together with source fragmentation.

We then proceeded to analyze a more complex mixture of free fatty acids, including mono- and polyunsaturated species. A mixture of eleven fatty acids was split in two, labeled with either methylamine NeuCode label to give the methyl amides (**Figure 3.1.3 A)**, and then recombined at a 1:1 ratio. All of the fatty acids separated chromatographically and their measured ratios were all near the expected 1:1 ratio (**Figure 3.1.3 B**, **Table 3.1.1**).

Our NeuCode method allows for the analysis of two samples per LC-MS experiment, with up to four samples per hour. Optimization of the LC gradient could undoubtedly increase throughput as could adding more NeuCode channels. Deuterium can be incorporated into methylamine labels to increase the number of channels up to four but could introduce chromatographic shifts. Adding ^{13}C and ^{15}N atoms is preferable to deuterium, but requires a larger label such as cholamine.[67] Cholamine offers three NeuCode channels, each spaced 6.32 mDa apart ($^{15}N_2$, $^{15}N^{13}C$, and $^{13}C_2$). Use of ^{13}C in the three methyl groups on the quaternary ammonium group of cholamine would double capacity to six by having two clusters of three peaks spaced 3 Da apart from each other.

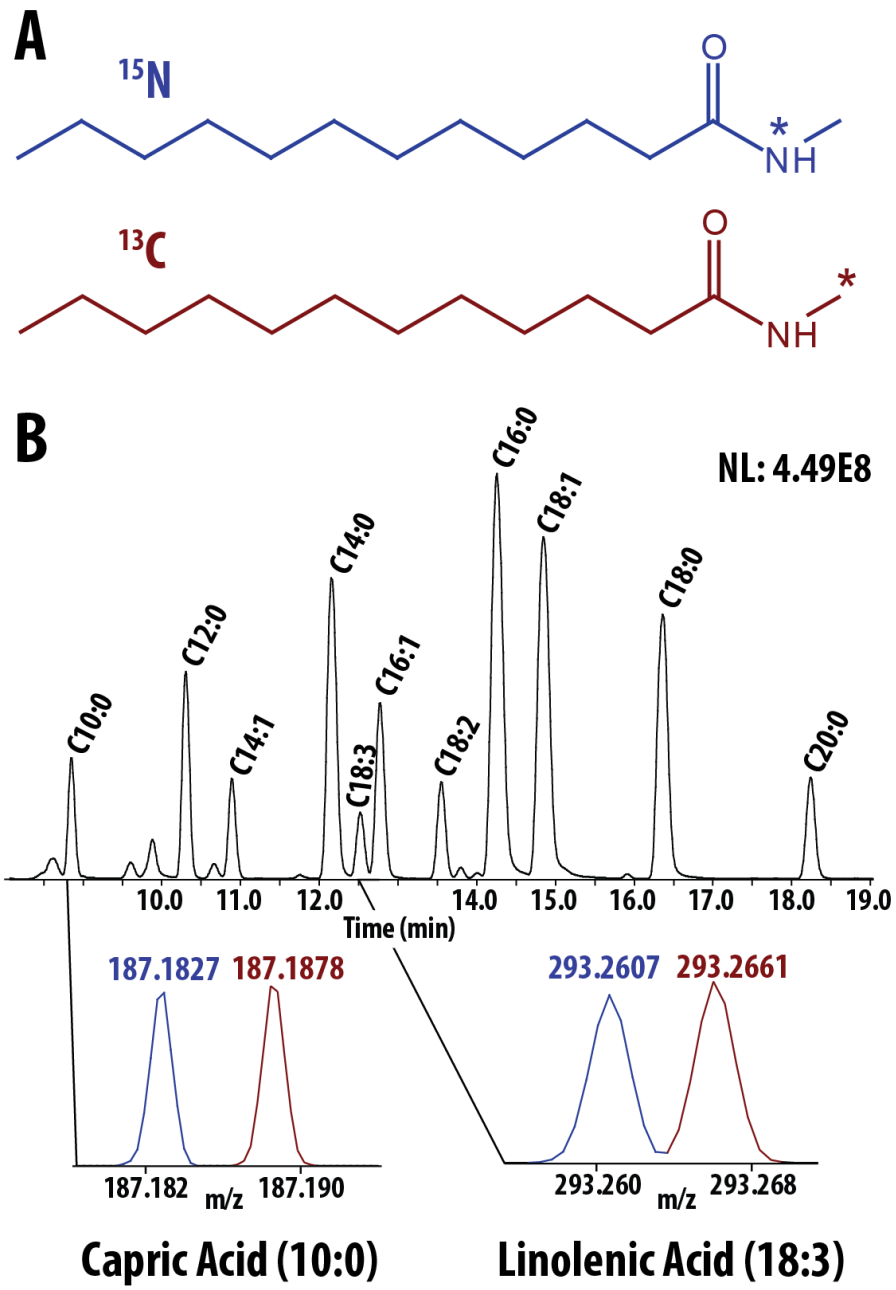
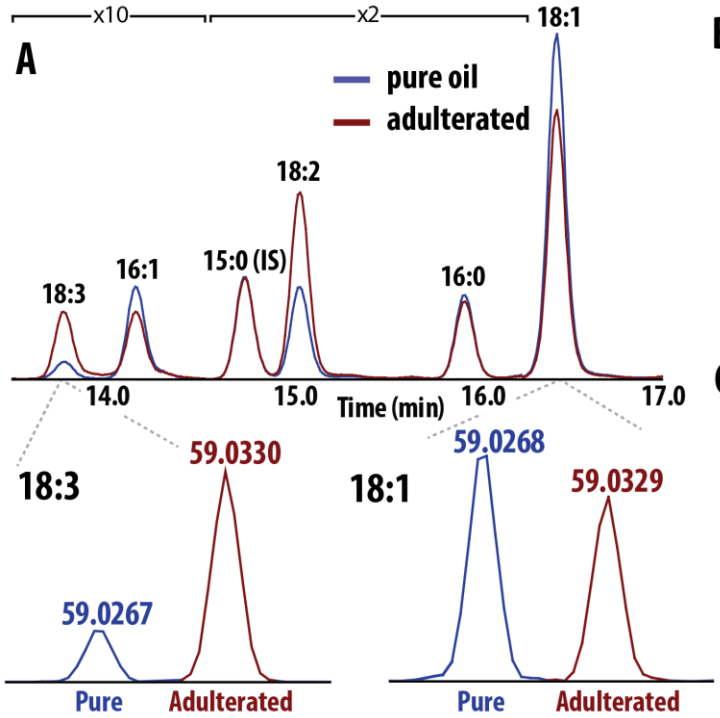


Figure 3.1.3. Labeling and quantitation of fatty acids via their methyl amides. **A** Structure of a fatty acid methyl amide and positions of the introduced isotopes. **B** Base peak chromatogram of fatty acid methyl amides prepared from a mix of free fatty acids from capric (C10:0) through arachidic (C20:0) acid. Quantitative molecular ion peaks from capric and linolenic acid at 140,000 RP are shown below the chromatogram.

Calibration curves from fatty acid methyl amides were prepared to determine linear dynamic range and to confirm suitability of the methylamine system for routine absolute quantitation in addition to relative quantitation. They yielded limits of detection for the fatty acids found in olive oil (palmitic, palmitoleic, stearic, oleic, linoleic, linolenic, arachidic acid, full scan mode) between 0.1 and 0.4 ng / mL (500 fg – 2 pg on column) and limits of quantitation ranged from 0.4 to 1.6 ng / mL. Linear dynamic range was at least 200 during initial trials. Correlation coefficients for all calibration curves were at least 0.998. Note that operation of the instrument using source fragmentation and selected ion monitoring (SIM) mode at 57 - 63 *m/z* reduced the limits of detection by a factor of seven to ten, i.e. between 12 to 40 pg / mL. We investigated the effect of increasing resolving power on linear dynamic range of 15:0 methyl amide quantitation and found a decrease in the correlation coefficient from 0.999 at RP = 17,500 to 0.991 at RP =140,000. Note that high RP is not necessary for quantitation from the universal fragment at 59 *m/z*. No significant difference between quantitation from the source fragment or the $[M+H]^+$ ion was found, but conditions where a difference in accuracy occurs may exist. For example, for very narrow peaks the number of points across the peak at 140,000 RP may become insufficient.

Analysis of edible oils

One common analysis of fatty acids is quantifying their abundances in edible oils (e.g., olive, canola, soybean, etc.) to determine the oils' quality and authenticity. We presumed that NeuCode quantitation would be an ideal platform for this type of analysis as two samples can be



B

fatty acid	Olive Oil	Soybean Oil
18:3	0.0%	5.9%
16:1	1.8%	0.0%
18:2	12.4%	54.3%
16:0	14.0%	7.9%
18:1	69.9%	17.8%
18:0	1.8%	3.0%

C

fatty acid	Predicted adult. : pure	Measured adult. : pure
18:3	∞	4.24
16:1	0.70	0.80
18:2	2.01	2.01
16:0	0.87	0.93
18:1	0.78	0.76
18:0	1.20	1.36

Figure 3.1.4. Determination of adulteration of olive oil with soybean oil. **A** Pure olive oil and adulterated olive oil base peak chromatograms of the fragment ions at 59.027 and 59.033 *m/z*. Total material on column was ~ 25 ng. A clear increase in linoleic to oleic acid ratio upon adulteration is visible. *M/z* peaks of the NeuCode ions of linolenic acid and oleic acid are shown below the. **B** Relative abundances of fatty acids in olive oil and soybean oil as measured by GC-MS. **C** The predicted ratio of each acid's abundance (GC-MS analysis) between the adulterated and pure samples and the actually measured ratio (NeuCode LC-MS analysis).

fatty acid	measured ratio
10:0	1.07
12:0	1.06
14:1	1.09
14:0	1.09
18:3	1.07
16:1	1.09
18:2	1.07
16:0	1.08
18:1	1.09
18:0	1.10
20:0	1.10

Table 3.1.1. The measured ratios of the quantitative NeuCode ions of fatty acid methyl amides. The samples were mixed at a 1:1 ratio and relative quantification proceeded from the elution profiles of the NeuCode ions at 59.027 and 59.033 *m/z*.

analyzed in tandem. To demonstrate NeuCode's applicability, we adulterated an Italian olive oil blend by adding 30% soybean oil. The pure olive oil was labeled with light (^{15}N) methylamine and the adulterated sample was labeled with heavy (^{13}C) methylamine. The two samples (unadulterated and adulterated) were then mixed 1:1 and analyzed by HPLC-MS as described above (**Figure 3.1.4**). Adulteration was detected by the change in relative abundance of fatty acid side chains upon addition of soybean oil. Addition of small amounts of soybean oil, which is rich in the polyunsaturated linoleic and linolenic acids, significantly increases their relative abundance in olive oil, which contains much smaller amounts of these fatty acids.[75] **Figure 3.1.4 A** shows the base peak chromatograms of the heavy and light NeuCode ions with the adulteration clearly detectable. This is evident by the increase in linoleic and linolenic acid abundance and by the concomitant decrease in the ratio of oleic to linoleic acid. The fatty acid profiles of the olive and soybean oil were each determined by GC-MS and changes in fatty acid abundance after adulteration with 30% soybean oil were predicted mathematically (**Figure 3.1.4 B**, **3.1.4C, eq. 3.1.2**). For example, relative abundance of linoleic acid was determined to be 12.4% in olive oil and 54.3% in soybean oil. Based on these GC-MS results, the predicted change in relative abundance of linoleic acid upon adulteration of olive oil with 30% soybean oil was calculated by:

$$\Delta\, linoleic\ acid = \frac{(0.7 * 0.124) + (0.3 * 0.543)}{0.124} = 2.01 \quad \text{eq. 3.1.2}$$

The measured fold-change by NeuCode LC-MS was 2.01. Other predicted decreases or increases in abundance of each fatty acid upon adulteration are closely approximated by the NeuCode method as well (**Figure 3.1.4 C**). Due to the poorer sensitivity of our GC-MS method no ratio could be predicted for linolenic acid or longer chain fatty acids.

Conclusions

We have demonstrated that methylamine NeuCode labels allow for positive mode LC-MS quantitation of fatty acid abundance in oils via the corresponding methyl amides, which can be detected with good sensitivity to below 1 ng / mL (1 ppb). Organic acids other than fatty acids can also be efficiently labeled with methylamine to give the methyl amides. NeuCode labeling doubles throughput by the concurrent analysis of two labeled samples. In another experiment, adulteration of olive oil with 30% soybean oil was easily detected in one LC-MS analysis. Resolving power requirements for organic acid analysis using NeuCode are lower than for NeuCode peptide analysis because ions bear only a single charge and are generally found at lower *m/z* values. Resolution requirements are further relaxed, when source fragmentation, all ion fragmentation, or MS/MS is used, since fragment ions are found at much lower *m/z* values while the spacing between the quantitative ions remains constant and, thus, analysis of NeuCode labeled samples on TOF systems is feasible. Finally, this quantitation method can be expanded by using a tag that contains more than one carbon or nitrogen atom. For example, six cholamine isotopologues could give rise to two clusters of quantitative triplets spaced three Dalton apart allowing for simultaneous 6-plex analysis.

Acknowledgments

We gratefully acknowledge Alan Higbee for help with GC-MS analysis of oil samples and the DOE Great Lakes Bioenergy Research Center (DE-FC02-07ER64494) for providing financial support.

Supporting Information Available

An Excel spreadsheet of theoretical resolving power calculations and MathCad spreadsheets used for peak modeling are available online through the American Chemical society at: http://pubs.acs.org/doi/abs/10.1021/ac500270q

3.2: NeuCode-Enabled Duplex Quantitation of Fatty Acid Methyl Esters by GC-MS

Abstract

A simple and cost-effective method for doubling the throughput of fatty acid methyl ester (FAME) analysis by GC-MS is described in this chapter. Sample preparation is simple and identical to existing standard methods. Instead of using regular, “light” methanol we use isotopically labeled methanol containing wither one ^{13}C or one ^{2}H atom. This encodes a 2.9 mDa mass difference between differentially labelled FAMEs that adds a second quantitative channel when a high resolution mass spectrometer is used. The method is especially well-suited for tests of identity and consistency of oils.

Introduction

The previous chapter and the preceding section of this chapter described methods using NeuCode quantitation for proteomics and for LC-MS based analysis of small molecules, an extension of other NeuCode proteomics work done in the Coon Research Group.[16,17,22,69,71] The NeuCode technique is not confined to LC-based separations, however. This chapter describes the further extension of the method to GC-MS analysis, rounding out the technique as truly versatile. The common assay of fatty acid methyl esters (FAMEs) by GC-MS was improved by doubling throughput and allowing more direct comparisons of samples by using ^{13}C and ^{2}H labeled heavy methanol for the preparation of the FAMEs. This gives rise to two quantitation channels from the ions containing the functional group. The characteristic ion at mass 74 (75 when labeled) arising from McLafferty rearrangement ion is NeuCode labeled for example, as are the ions used for quantitation of unsaturated fatty acids. This work presents an affordable and straightforward way to multiplex FAME quantitation that can be conducted on GC-Orbitrap and GC-ToF instruments. Despite the use of isotopically labeled reagents, the total reagent cost per pair of samples is only about $6. The official methods put forth by the American Oil

Chemists' Society (AOCS) for the preparation of FAMEs can be used without modification, allowing non-expert technicians to use the NeuCode method.[72] Finally, the use of higher alcohols, such as ethanol or isopropanol, would allow the expansion of the method to more than two channels.

Experimental

Materials

Deuterium labeled methanol was obtained from CDN Isotopes (Pointe-Claire, QC, Canada) at 98% enrichment and ^{13}C labeled methanol was obtained from Sigma Aldrich (St. Louis, MO) at 99% enrichment. Olive Oil was obtained from HyVee (Classic Select Olive Oil, Des Moines, IA), coconut oil from Trader Joe's (Organic Virgin Coconut Oil, Monrovia, CA), and soybean oil from The J.M. Smucker Company (Crisco Vegetable Oil, Orrville, OH). Anhydrous hexane (Sigma Aldrich, St. Louis, MO), and KOH pellets (Thermo Fisher Scientific, Waltham, MA) were used without further purification. Deionized water was prepared in house (NANOpure Diamond, Barnstead, Thermo Fisher Scientific, Waltham, MA). The analysis of samples took place on a Trace GC (Thermo Fisher Scientific) equipped with a 5% diphenyl non-polar capillary column(Rxi-5Sil, 30m x 0.25 mm, Restek, Bellefonte, PA) column coupled to a modified Q Exactive mass spectrometer (Thermo Fisher Scientific).[76]

Methods

FAMEs were prepared from the oils following the recommended alternate AOCS method Ce 2-66.[72] Equimolar portions of the oils (99 – 110 mg) were weighed into 2 mL screw-cap glass vials. Then, 1.00 mL hexanes were added followed by 50 µL of 2N KOH in heavy methanol. The vials were capped and vortexed for 30 s. **Figure 3.2.1** shows the reaction and the resulting differentially labeled fatty acids. After centrifuging, 2 µL of the upper layers were removed with a micropipette and added to 1.00 mL hexanes. This solution was analyzed by GC-MS. A 1:10 split

injection of 1 µL of the solution was made and the FAMEs were separated by a non-linear temperature gradient from 80˚C to 310˚C on a non-polar (DB-5) capillary GC column. The temperature was held at 80 ˚C for 1 min, then raised to 195 ˚C at 50 ˚C/min, then ramped to 212 ˚C at 2 ˚C/min, then ramped to 310 ˚C/min at 15 ˚C/min and held there for 90 seconds. The data were recorded at a resolving power setting of 35,000 with an AGC setting of 5e5 charges and the data were analyzed using the Xcalibur software suite (Thermo Fisher Scientific, San Jose, CA).

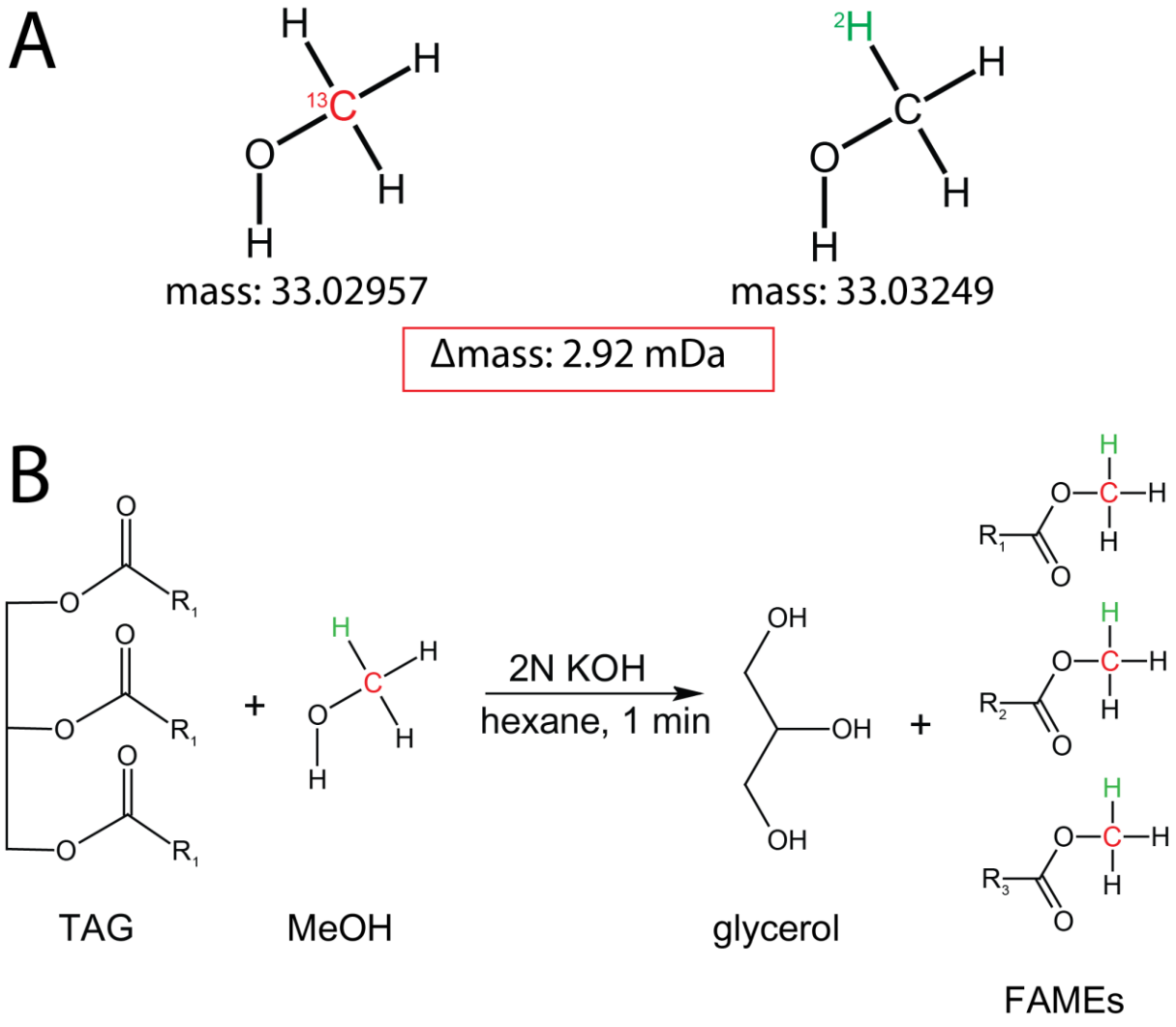


Figure 3.2.1. Synthesis of isotopically labeled FAMEs. **A** Deuterium and ^{13}C labeled methanol is shown with the two species differing in mass by 2.9 mDa. **B** The standard procedure for

synthesizing FAMEs is shown. The triacylglycerides are transesterified in strong base to the methyl esters.

Results

The NeuCode peak pairs spaced at 3 mTh can be clearly observed in the acquired spectra and their intensities reflect the relative concentration of each fatty acid in the sample. All fragments in the mass spectrum that contain the methyl moiety give raise to two closely spaced peaks that can be used for quantitation. For saturated fatty acids, this includes the McLafferty ion at *m/z* 75 (shifted up one mass unit due to the NeuCode label) and the acrylate-type fragment at *m/z* 88 (**Figure 3.2.2 C**).[11] Unsaturated species also give rise to quantitative ions that are analogous to their unlabeled equivalents shifted up one mass unit. Two samples of olive oil were differentially labeled and then re-combined in a 1:1 ratio to successfully confirm the quantitative nature of NeuCode FAME synthesis (**Figure 3.2.2 A-C**). The technique can also be used to establish authenticity of an oil or to monitor variation between samples. This application would be useful in a forensic environment and in an industrial environment were product consistency is very important. We mixed equimolar samples of coconut and olive oil and saw the NeuCode peak ratios reflect the drastically different fatty acid composition of the two oils. For lower carbon number saturated fatty acids (8 – 12 carbon atoms), only the peak from the coconut channel was present, whereas for long chain unsaturated acids only the olive oil channel was present. Some fatty acids are present in both oils in appreciable quantities and thus peak pairs of varying ratios were observed (16:0, 18:0, and 20:0) (**Figure 3.2.2 C**).

Conclusion

The adaptation of NeuCode to GC-MS analysis of FAMEs provides a straightforward and cost-effective way to multiplex analyses, thereby saving instrument time and providing a more immediate comparison between two samples. This comparison could be used routinely without the need for a standard curve to test samples for identity or adulteration or to monitor product variance over time.

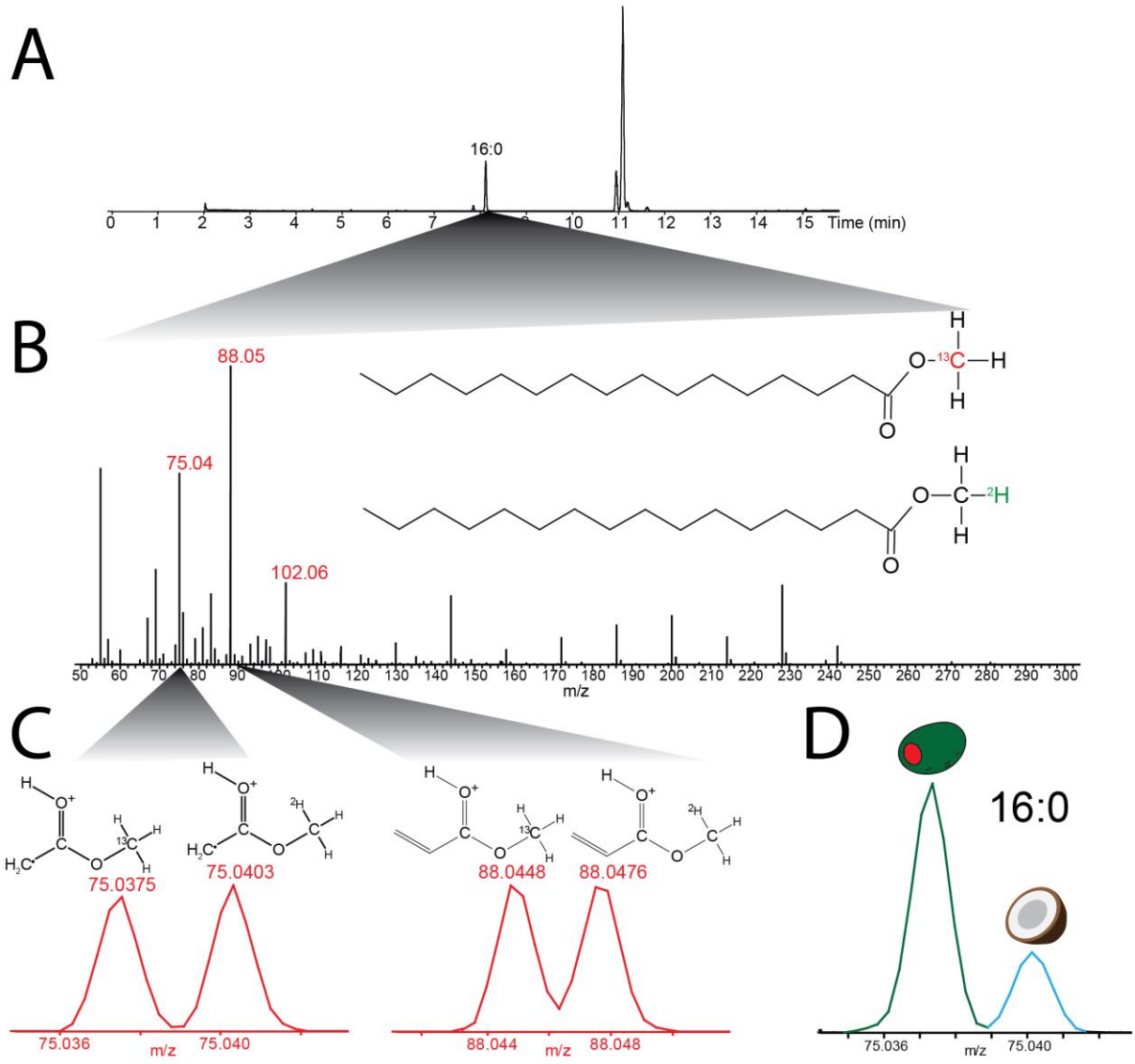


Figure 3.2.2. Application of FAME NeuCode to compare two samples of olive oil and olive oil to coconut oil. **A** The total ion chromatogram of an injection of two samples of olive oil at a 1:1 ration. **B** The EI mass spectrum of palmitic acid methyl ester containing ^{13}C or ^{2}H. Peaks containing the NeuCode label are shifted up one unit in mass and peaks suitable for quantitation at moderate resolving powers are labeled in red. **C** Two NeuCode peak pairs are enlarged to reveal the quantitative information. Their abundance accurately reflects the 1:1 mixing ratio of the two samples. **D** The palmitic acid content of samples of olive oil and coconut oil mixed 1:1. Coconut oil contains less since it consists chiefly of shorter chain saturated fatty acids.

Chapter 4: Segmentation of precursor mass range using 'tiling' approach increases peptide identifications for MS1-based label-free quantification

A.U. designed experiments, conducted experiments, and wrote the paper together with GKP and CEM.

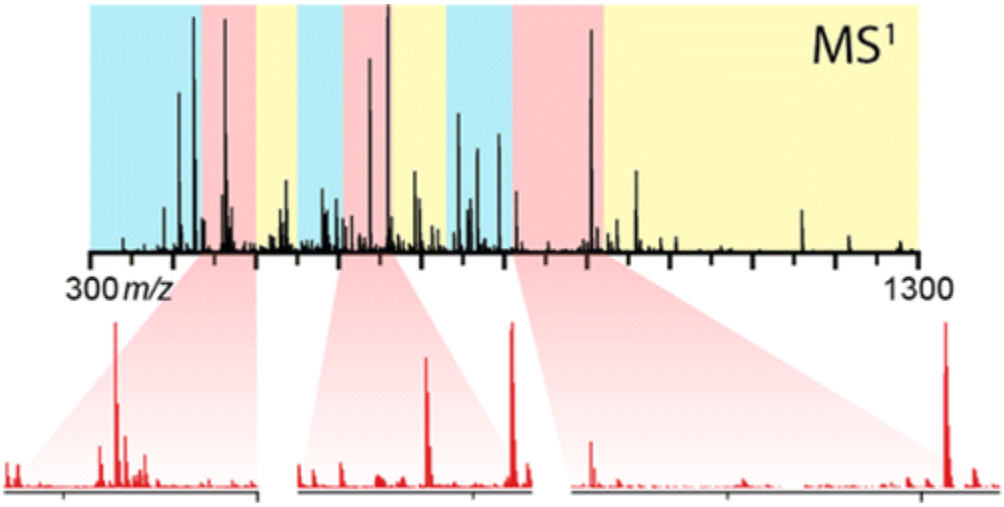
This chapter has been published:

*Vincent, C. E., *Potts, G. K., *Ulbrich, A., Westphall, M. S., Atwood III, J. A., Coon, J. J., & Weatherly, D. B. (2013). Segmentation of precursor mass range using "tiling" approach increases peptide identifications for MS1-based label-free quantification. *Analytical chemistry*, *85*(5), 2825-2832.

* co-first authors

Abstract

Label-free quantification is a powerful tool for the measurement of protein abundances by mass spectrometric methods. To maximize quantifiable identifications, MS^1-based methods must balance the collection of survey scans and fragmentation spectra while maintaining reproducible extracted ion chromatograms (XIC). Here we present a method which increases the depth of proteome coverage over replicate data-dependent experiments without the requirement of additional instrument time or sample pre-fractionation. Sampling depth is increased by restricting precursor selection to a fraction of the full MS^1 mass range for each replicate; collectively, the *m/z* segments of all replicates encompass the full MS^1 range. Although selection windows are narrowed, full MS^1 spectra are obtained throughout the method, enabling the collection of full mass range MS^1 chromatograms such that label-free quantitation can be performed for any peptide in any experiment. We term this approach “binning” or “tiling” depending on the type of *m/z* window utilized. By combining the data obtained from each segment, we find that this approach increases the number of quantifiable yeast peptides and proteins by 31% and 52%, respectively, when compared to normal data-dependent experiments performed in replicate.



Introduction

Mass spectrometry methods for quantitative proteomics aim to maximize protein identifications and accurately characterize protein abundance in a cost- and time-efficient manner. MS^1-based label free methods are an attractive option for relative protein quantification, as they eliminate the expenses and sample preparation associated with isotope or mass tag labeling techniques.[45,77,78] To attain quantitative data, these methods exploit the linear relationship of peptide spectral peak intensity and relative peptide abundance in a mixture.[78-81] Each sample individually undergoes LC-MS/MS analysis, and extracted ion chromatogram (XIC) signal intensities from identical peptides are then compared across the separate analyses, such that the relative abundance of their parent proteins within the different samples can be determined.[82-84] The development of algorithms to facilitate chromatogram alignment has been crucial for these XIC comparisons, but highly reproducible separations remain essential for the acquisition of reliable quantitative data using an MS^1-based approach.[85-89] As in other proteomics experiments, the maximization of protein identifications using MS^1-based label-free methods becomes more daunting as sample complexity increases. Traditional data-dependent acquisition favors the highest-intensity peptides for analysis, which can preclude the sampling and identification of species present at low signal-to-noise. The reduction of sample complexity afforded by off-line fractionation facilitates an increase in attainable peptide identifications.[53,90-94] The creation of multiple fractions from one complex sample disperses high-abundance peptides over multiple experiments, enabling the detection, sampling, and identification of less abundant species from reduced MS^1 complexity spectra. Unfortunately, overall sample loss and variable peptide elution across fractions are inevitable consequences of off-line fractionation, and these effects introduce additional challenges to chromatogram alignment for label-free MS^1-based quantification.[34] Various post-acquisition data analysis strategies have been developed to correct for any systematic bias off-line fractionation introduces to MS^1-based label free

analyses,[34,95] but the ability to increase identifications without having to devote extra time to sample preparation and post-acquisition analysis would be advantageous.

Online fractionation techniques, and the use of longer chromatography columns and/or extended chromatographic gradients, also reduce sample complexity at the mass spectrometer, increasing overall peptide identifications.[96-98] These methods improve chromatographic resolution over an extended LC-MS/MS analysis time to boost the number of peptides that will be detected, sampled, and identified during a single LC-MS/MS experiment. Another alternative to off-line fractionation is gas-phase fractionation (GPF) in the mass-to-charge (*m/z*) dimension.[35,54,99,100] In the GPF technique, MS^1 or SIM scans are performed over a narrow *m/z* range (e.g. 600 – 700 instead of the full 300 – 1300 mass space). Precursor selection is restricted to this truncated scan range, which allows extension of sampling depth into lower intensity features within the *m/z* region. By interrogating sub-sections of the MS^1 mass range in sequential experiments, GPF increases peptide identifications compared to normal data-dependent methods.[100] The aforementioned strategies are useful alternatives to off-line fractionation for the maximization of peptide identifications in MS^1-based label-free analyses, since the ability to inject and analyze an unfractionated sample for each experiment limits run-to-run chromatographic variability that could compromise quantification. The major drawback of fractionation and longer gradients, however, is increased time for analysis. Every sample in an MS^1-based label-free study requires a separate LC-MS/MS experiment, and, as the number of samples involved in the comparison and/or the number of replicate experiments to perform increases, the total cost associated with analysis time also increases, compared to traditional data-dependent acquisition.

Here we report a strategy which attains a greater number of quantifiable peptide identifications than replicate data-dependent experiments, without the requirement of off-line fractionation or additional instrument analysis time. To achieve greater sampling depth than traditional data-

dependent acquisition, we narrow the mass range from which a precursor can be selected for MS^2 analysis. This idea is similar to that employed for GPF, except the MS^1 scans, from which precursors are selected, encompass the full mass range (in our case, *m/z* 300 – 1300). Scherl et. al. employed a similar strategy when they limited precursor selection to truncated, continuous regions of the MS^1 mass range; their study found that the depth of proteome coverage was increased as a result.[101] We refer to this method as 'binning.' We demonstrate that, by restricting precursor selection to several narrowed *m/z* ranges distributed throughout the full MS^1 mass range ('tiling') instead of one continuous region ('binning'), we can boost peptide identifications to an even greater extent, without the requirement of additional instrument analysis time. The 'binning' and 'tiling' methods are ideal for the maximization of protein identifications in MS^1-based label-free analyses, as they maintain acquisition of the full MS^1 mass range throughout all experiments. This not only preserves high chromatographic reproducibility between LC-MS/MS analyses for accurate XIC comparisons, but it also enables the acquisition of replicate data within one set of experiments (a peptide identified in one experiment should theoretically be quantifiable in all 'binning' or 'tiling' experiments). We apply our 'binning' and 'tiling' approaches to the MS^1-based label-free quantitative analysis of a complex mixture of tryptically-derived yeast peptides and demonstrate: 1) a significant boost in quantifiable identifications over traditional data-dependent acquisition (DDA); and 2) a reduction in analysis time, compared to the acquisition of replicate DDA analyses of off-line fractionated samples.

Materials and Methods

Cell culture, differentiation, and lysis

Wild-type yeast (*Saccharomyces cerevisiae*) was grown in rich medium to an OD600 of 0.6. Cells were collected and centrifuged at 14,200 g for 10 min at 4 °C. The resulting cell pellet was washed twice with sterile water and centrifuged at 1,100 g for 5 min. Lysis buffer of

approximately three times the cell pellet volume was added. The lysis buffer contained 8 M urea, 75 mM NaCl, 50 mM Tris (pH 8.2), 1 mM sodium orthovanadate, 100 mM sodium butyrate, complete mini EDTA-free protease inhibitor (Roche Diagnostics, Indianapolis, IN) and phosSTOP phosphatase inhibitor (Roche Diagnostics, Indianapolis, IN). Yeast cells were frozen into small droplets using liquid nitrogen. The resulting pellets were placed in a pre-chilled grinder jar (Restek, Bellefonte, PA) in equal volume with acid-washed glass beads (Sigma Aldrich, St. Louis, MO). The cells were lysed using a MM4000 Mixer Mill (Restek, Bellefonte, PA), where the sample was beat at 30 Hz for 4 minutes a total of three times. The lysate was centrifuged for 10 minutes at 4,000 rpm at 4°C.

Cell digestion

Yeast proteins were subjected to cysteine residue reduction using 5 mM DTT and alkylation using 10 mM iodoacetamide. The protein sample was diluted to a final concentration of 1.5 M urea (pH 8) with a solution of 25 mM Tris and 2 mM $CaCl_2$. Sequencing-grade trypsin (Promega, Madison, WI) was added to each sample at a ratio of 1:50 (enzyme:protein) and the resulting mixture was incubated at 37°C overnight. The reaction was quenched using trifluoroacetic acid. The sample was desalted using C18 solid-phase extraction columns (SepPak, Waters, Milford, MA) and dried to completion).

Sample Preparation

Yeast peptides were re-suspended in 0.2% formic acid and split into two equal mass aliquots. One aliquot was used for LC-MS analysis (unfractionated sample). The other aliquot was subjected to reverse-phase (RP) fractionation. The digested peptides were dried to completion and dissolved in 100 µL of RP buffer A [20 mM NH_4HCO_2 (pH 10)]. The sample was injected onto a column packed with non-polar material (Gemini 5 µm C18 110 Å LC Column 250 × 4.6 mm, Phenomenex, Torrance, CA) attached to a Surveyor LC quaternary pump (Thermo

Electron, West Chester, PA) running at 3.0 mL/min. Peptides were detected using a PDA detector (Thermo Electron, West Chester, PA). The eluate was collected in 30-sec intervals starting 12 minutes into the following gradient: 4 min of 95% RP buffer A and 5% RP buffer B [20 mM NH_4HCO_2, 80% ACN (pH 10)], followed by a linear gradient of 5–12% RP buffer B from 4 to 8 min, followed by a linear gradient of 12–45% RP buffer B from 8 to 34 min, followed by a linear gradient of 45% – 100% RP buffer B from 34 to 36 min. RP buffer B was held at 100% for 5 minutes, after which there was a 1 min transition to 5% RP buffer B. The column was re-equilibrated at 5% RP buffer B for 20 minutes. A total of 40 fractions were collected into 20 vials (each vial contained two 30-sec fractions collected 10 minutes apart). The 20 vials were pooled into 5 samples such that each sample contained fractions collected 2.5 minutes apart. All samples were dried on a vacuum centrifuge and resuspended in 0.2% formic acid for LC-MS analysis.

Liquid chromatography-mass spectrometry

All experiments were performed using a NanoAcquity UPLC system (Waters, Milford, MA) coupled to a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Samples were loaded onto a precolumn (75 µm i.d., packed with 10 cm of 5 µm C18 particles, pore size 100Å; Microm Bioresources Inc, Auburn, CA) for 10 minutes at 98:2 buffer A [0.2% formic acid]:buffer B [acetonitrile with 0.2% formic acid] at a flow rate of 1.0 µL/min. Samples were then separated on an analytical column (75 µm i.d., packed with 15 cm of 5 µm C18 particles, pore size 100Å; Microm Bioresources Inc, Auburn, CA) at a flow rate of 0.300 µL/min using the following gradient: an initial steep rise to 8% B in 1 min, followed by a 59 min linear gradient from 8% to 30% B, followed by a final 5 min ramp to 70% B which was held for 5 minutes. The column was equilibrated with 2% buffer B for an additional 20 min. Precursor peptide cations were generated from the eluent through the utilization of a nanoESI source.

All instrument methods consisted of an MS^1 scan (300 – 1300 *m/z*) analyzed in the Orbitrap at a resolution (*m/Δm*) of 60,000 followed by ten data-dependent HCD MS^2 scans analyzed in the Orbitrap at a resolution (*m/Δm*) of 15,000. For the DDA control methods, precursors were selected from the entire MS^1 mass range for MS^2 analysis. For the binning and tiling methods, precursors were only selected for MS^2 analysis if they fell within specific *m/z* range(s). The *m/z* ranges designated for each experiment in this study are presented in Table 4.1. Restricted precursor selection for the tiling and binning techniques was implemented through slight modifications to the instrument control language (ITCL, access granted by Thermo Fisher Scientific). Note that while the tiling method requires ITCL modification, the binning method can be implemented using existing Xcalibur software (*m/z* range restriction set in the method file under data dependent settings –> global –> global –> mass range for selecting MS dependent masses).

All MS^2 scans employed a precursor isolation window of 2 Th and an HCD normalized collision energy (NCE) setting of 30 for 0.1 ms. The automatic gain control (AGC) target settings for precursor cations were 1×10^6 charges for MS^1 scans and 5×10^4 charges for HCD-activated MS^2 scans. Precursors were subject to dynamic exclusion for 30 seconds using a 10 ppm window.

Binning (3)	
Experiment	m/z ranges for precursor selection
1	300-552
2	552-730
3	730-1300
Tiling (3)	
Experiment	m/z ranges for precursor selection
1	300-435, 550-605, 730-810
2	435-500, 605-665, 810-920
3	500-550, 665-730, 920-1300
Tiling (5)	
Experiment	m/z ranges for precursor selection
1	300–367, 635–702, 970–1036
2	367–434, 702–769, 1036–1102
3	434–501, 769–836, 1102–1168
4	501–568, 836–903, 1168–1234
5	568–635, 903–970, 1234–1300

Table 4.1 MS^1 *m/z* ranges from which precursors were selected for MS^2 analysis in 'binning' and 'tiling' methods.

Data Analysis

Data was processed using the in-house software suite COMPASS.[42] OMSSA (version 2.1.8) searches were performed against the International Protein Index (IPI: http://www.ebi.ac.uk/IPI/) target-decoy database comprised of yeast (*Saccharomyces* Genome Database, http://www.yeastgenome.org, February 3, 2011, "all" version including all systematically named open reading frames (ORFs), including verified, uncharacterized, and dubious ORFs and pseudogenes) proteins.[41] Searches were conducted using a mono-isotopic precursor mass tolerance of ±5.0 Da and a mono-isotopic product mass tolerance of ±0.01 Da. The fixed modification specified was carbamidomethylation of cysteine residues and the variable modification specified was the oxidation of methionine residues. A maximum of 3 missed tryptic cleavages were allowed. For each method, data was acquired over multiple experiments and peptide identifications were collectively filtered to both 1% FDR and 10% FDR.

Comparisons between approaches were performed using a customized version (based on version 2.6) of the ProteoIQ software (NuSep Inc., Bogart, GA, www.nusep.com). Peptides from the filtered COMPASS results were loaded into ProteoIQ, matched to their corresponding proteins (using same yeast database), and organized into protein groups. Proteins were regarded as "present" in a sample if it was identified in the 1% FDR list and "absent" if it was not identified in the 10% FDR list.

Quantifiable peptides were defined as those that gave a "quant score" of at least 0.5, as determined by the ProteoIQ software. The quant score was above 0.5 when at least one observed isotopic distribution in a full MS^1 scan matched the theoretical isotopic distribution of the associated peptide identification with a product R^2 value of 0.5 (the product of the R^2-values based on observed vs. theoretical relative intensity and observed vs. theoretical mass accuracy).

Results and Discussion

Here we present a method that increases the depth of proteome coverage over replicate data-dependent experiments without the need for additional instrument analysis time or sample fractionation. To facilitate straightforward compatibility with MS^1-based label-free quantification, we designed this method to gather full MS^1 scans throughout each LC-MS/MS analysis, thereby limiting chromatographic variability and minimizing the time required for the acquisition of quantitative data in replicate.

Figure 4.1 A illustrates three different strategies for precursor selection using a full MS^1 scan. The first approach (**Figure 4.1 Ai**) is the traditional data-dependent acquisition (DDA), a strategy in which a user-defined number of the most intense precursors from the full MS^1 mass-to-charge (*m/z*) range are selected for MS^2 analysis. When this technique is utilized for the analysis of complex samples, the number of achievable peptide identifications is often limited by sampling; in other words, selection of the most intense precursors from the full MS^1 *m/z* space can preclude the sampling and identification of lower abundance species. The plots in **Figure 4.1 A** show the *m/z* distribution of precursors sampled during a DDA control run. The color of each point, and the bars on the right hand side of each panel, indicate the experiment in which each precursor would have been sampled within each of the respective methods. Note that, for visualization purposes, each point in the DDA plot was assigned a random color; any precursor can be selected in any DDA experiment, and many will be selected in more than one LC-MS/MS analysis.

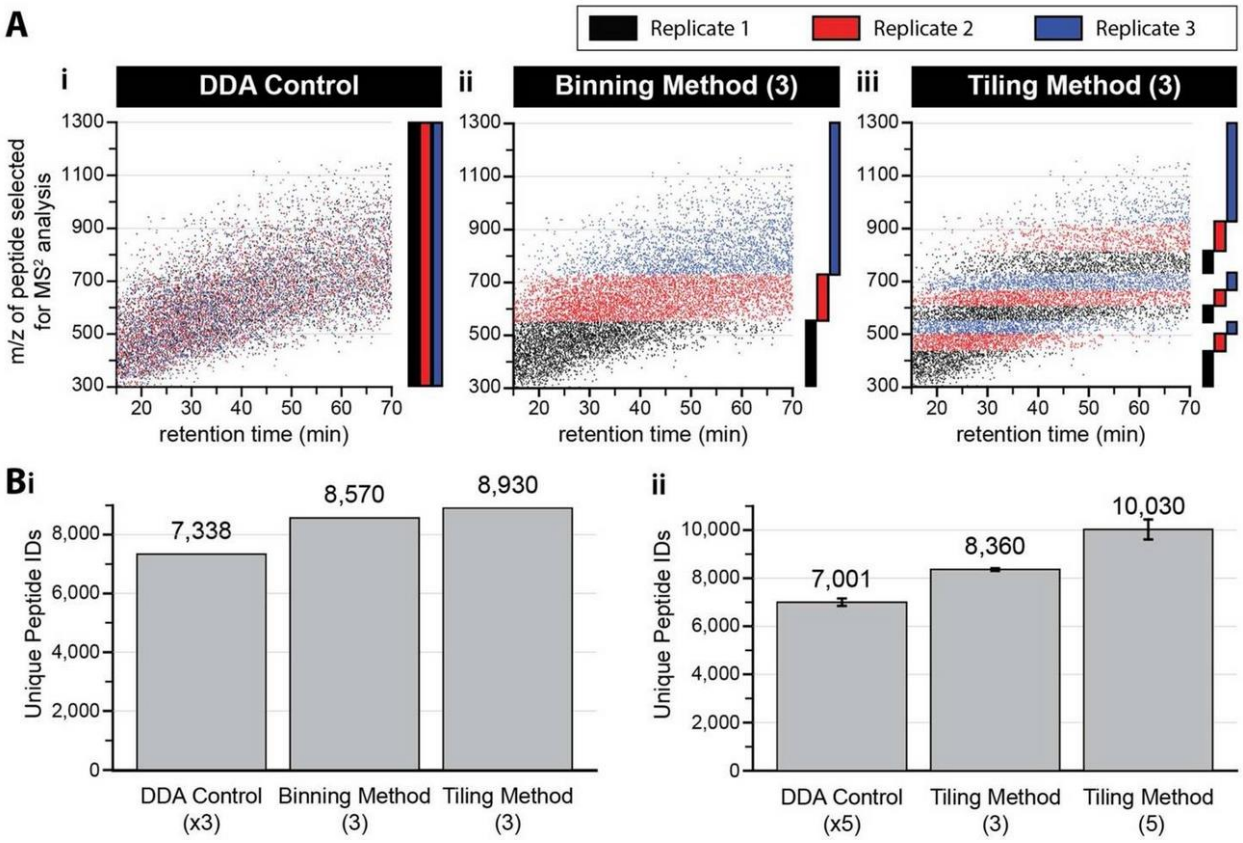


Figure 4.1 Tiling method achieves more unique peptide identifications than binning and data-dependent methods **A**. Mass-to-charge (*m/z*) values of precursors selected for MS^2 analysis during data-dependent acquisition over a 55 minute LC gradient. Any restrictions imposed for precursor *m/z* selection in the individual methods are represented by the red, blue, and black dots (in plots) and bars (to right of plots). **B**. Comparison of unique peptide identifications obtained using i) different precursor selection requirements and ii) a different number of 'tiles' (n = 2). Note that data for figures (i) and (ii) were obtained on different instruments on different dates; this accounts for identification discrepancies observed between the two figures.

Unlike DDA, the second and third strategies (**Figure 4.1 Aii–iii**) restrict precursor selection to designated segments of the full MS^1 mass range. In both methods, an unfractionated sample is subjected to a series of LC-MS/MS analyses (one for each *m/z* segment). Full MS^1 scans are collected throughout each analysis, and ions are only eligible for precursor selection if they fall inside the designated *m/z* segment of the MS^1 scan. By restricting precursor interrogation in this manner, the LC-MS/MS analyses will sample deeper into each MS^1 spectra, enabling the investigation of low-intensity peptides, which wouldn't typically be sampled by DDA. At the conclusion of the experimental sets, peptide identifications obtained from all analyses provide a comprehensive list of precursors sampled across the full MS^1 mass range.

We investigate two approaches for the creation of *m/z* segments from the full MS^1 mass range. The first strategy, 'binning,' divides the full MS^1 range into continuous, sequential segments (**Figure 4.1 Aii**). The second strategy, 'tiling,' partitions the full MS^1 range into segments which contain several small *m/z* sections spanning upper, middle, and lower *m/z* regions of the full MS^1 mass range (**Figure 4.1 Aiii**). The *m/z* range(s) assigned to each segment can be determined in several ways,[100,101] here we use two: 1) segments are created such that they span equal *m/z* ranges; and 2) boundaries are set such that each segment produces an equivalent number of peptide identifications. Note that to optimally implement this second approach, a control DDA run must be performed prior to the analysis so that the *m/z* distribution of peptides in the sample can be determined for the LC-MS method employed. The advantage to this second approach is that the most abundant peptides in the sample will be equally distributed between each of the segments.

Unique Peptide Identifications Achievable with Binning, Tiling, and Data-dependent *Methods*

The ability of the binning and tiling methods to improve peptide identification was assessed using a complex peptide mixture generated through digestion of yeast cell lysate with trypsin.

Binning and tiling methods were employed using three or five segments; the *m/z* boundaries of each segment were assigned based on the *m/z* distribution of precursors identified in a prior data-dependent run of the yeast sample. These methods were then compared to normal DDA performed in triplicate (control). **Figure 4.1** presents the outcome of these experiments. Using the binning and tiling strategies illustrated in **Figure 4.1 A**, we identified 17% and 22% more unique peptides, respectively, than were identified by three replicate data-dependent control methods (**Figure 4.1 Bi**).

As was observed by Scherl et. al.,[101] we find that precursor *m/z* generally increases with elution time (**Figure 4.1 A**). This indicates that, although each segment in the binning method contains an equal number of peptide precursors, these precursors are distributed unevenly along the LC gradient (**Figure 4.1 Aii**). This is detrimental to the acquisition of maximum peptide identifications. For example, the majority of the precursors sampled in replicate 1 (*m/z* 300 - 552) of **Figure 4.1 Aii** elute during the first third of the experiment. At this time, there are more precursors eluting in that *m/z* range than can be selected for MS/MS analysis, meaning peptides in this region are being under-sampled (**Suppl. Figure 4.1**). During the last third of this analysis, however, there are very few precursors eluting that fall within the acceptable *m/z* range. At this point in the gradient, there is excess time to sample all precursors available for MS/MS analysis, meaning sampling depth is maximized, but analysis time is not being effectively utilized. To reduce the sampling inefficiencies observed with the binning method, we developed a tiling scheme for the creation of *m/z* segments (**Figure 4.1 Aiii**). By incorporating low-, medium-, and high-*m/z* ranges within a single segment, the tiling scheme accounts for the changes in precursor *m/z* with elution time to ensure a more homogeneous distribution of peptide identifications across each experiment. This, presumably, led to the modest increase in peptide identifications achieved by the tiling method compared to the binning strategy, (8,930 vs. 8,570, respectively) (**Figure 4.1**, **Suppl. Figure 4.1**). Having found that the tiling method performs.

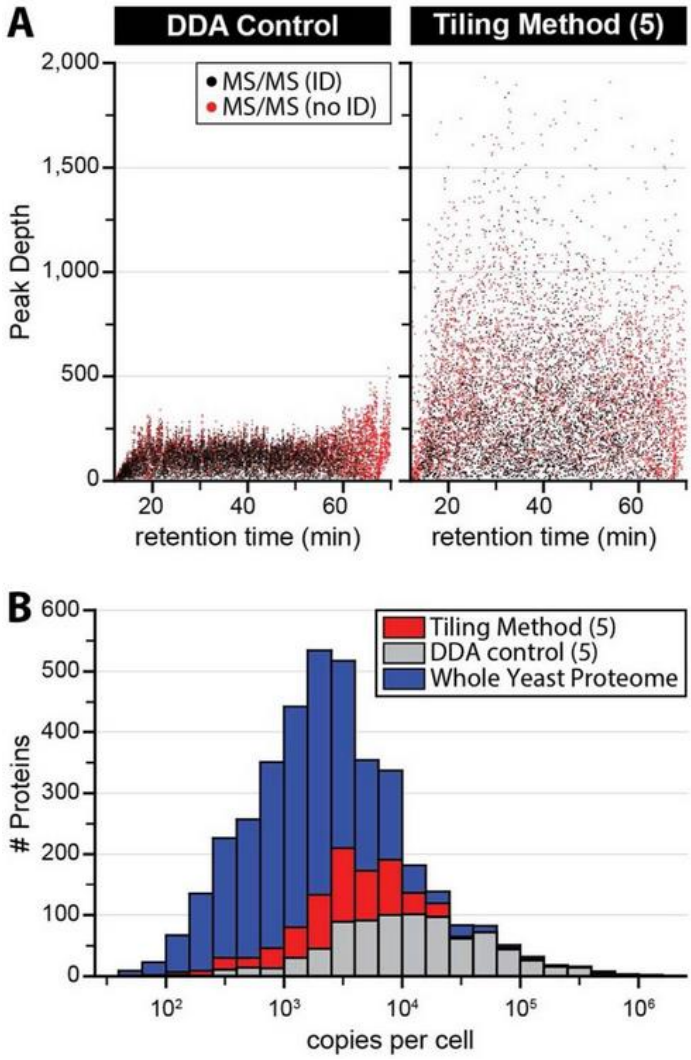


Figure 4.2 Tiling method enables the sampling and identification of low-abundance peptides which are not observable by regular data-dependent methods **A** Comparison of MS^1 peak depth for precursors selected and/or identified by DDA control vs. tiling methods over a 60 minute LC gradient. **B** Distribution of yeast proteins found in the yeast proteome (blue), the tiling experiments (red), and the DDA control experiments (grey) grouped based on relative abundance (protein copies per cell) within the yeast proteome.

just as well, and even slightly better, than the binning method, we chose to focus on the comparison between tiling and DDA methods for the remainder of our study

Crucial to the optimization of the tiling method is the balance between maximizing peptide identifications and minimizing required analysis time. Having only a few segments is time-efficient, but the sampling of low-level peptides is sacrificed, as higher intensity precursors within the large *m/z* segments are predominantly sampled. Conversely, increasing the total number of segments improves the probability of sampling low-level precursors, but incurs significant time costs; a greater number of segmented experiments will need to be performed, and time which could be utilized for precursor sampling will be allotted to acquiring MS^1 scans since too few precursors may exist in the small *m/z* segments to fill all ten data-dependent slots. By increasing the number of segments in our tiling method from three to five, we obtained a 20% boost in unique peptide identifications; this enabled the identification of 43% more unique peptides than could be obtained by five replicates of normal DDA (**Figure 4.1 Bii**).

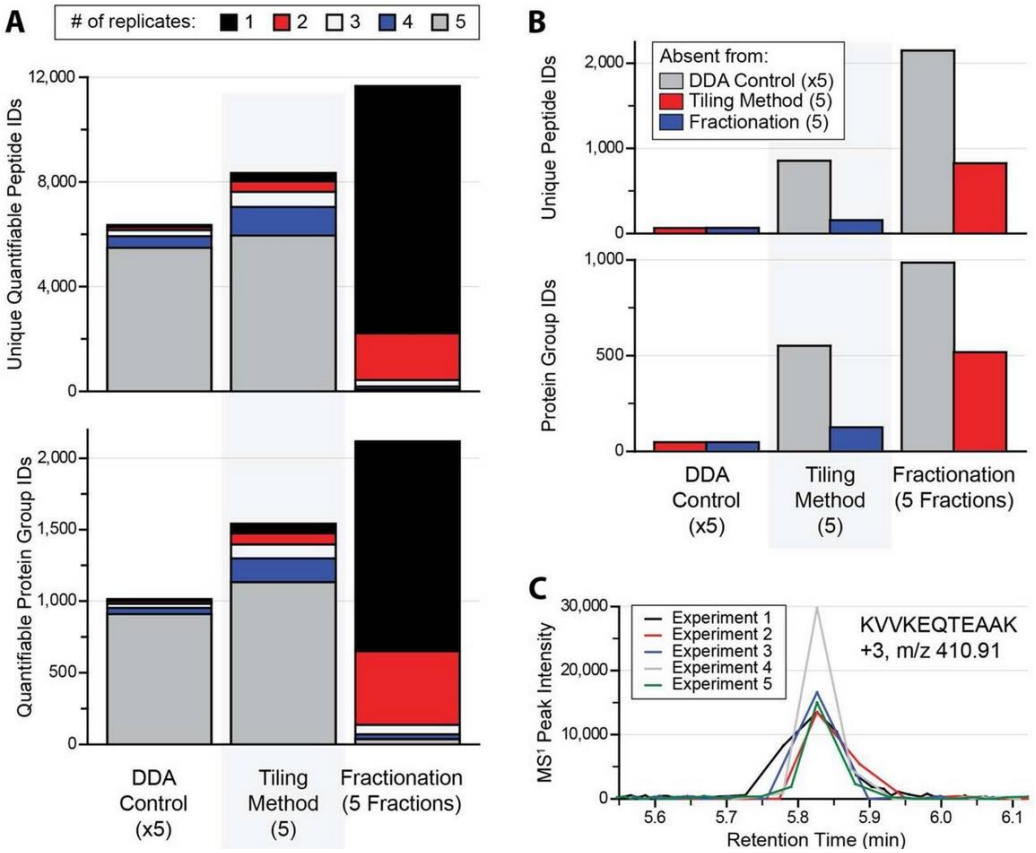


Figure 4.3 Tiling method identifies more quantifiable unique peptides and protein groups than regular data-dependent acquisition. Tiling method attains fever quantifiable identifications, but significantly more replicates, than fractionation over five LC-MS/MS analyses **A.** Quantifiable unique peptides and protein groups discovered in data-dependent acquisition (5 replicates), the 5-tiling method, and fractionation (5 fractions) performed on the same tryptic yeast digest. The number of replicate experiments in which each peptide or protein group could be quantified is depicted. **B.** Unique peptide and protein group identifications found in each of the three methods (at 1% FDR), but absent from each of the other methods (at 10% FDR). **C.** Extracted ion chromatograms for peptide KVVKEQTEAKK (*m/z* 410.91) from each of the 5-tiling experiments. This exemplifies one of many peptides which could be quantified in all five of the Tiling experiments, but could not be quantified in any of the data-dependent experiments.

Tiling Method Improves Sampling of Low Abundance Peptides and Proteins

By only sampling precursors from a certain *m/z* segment of the MS^1 spectrum, both binning and tiling methods increase the probability of sampling low intensity precursors that would often be passed over in normal data-dependent methods, in favor of more intense peptides with *m/z* values outside of the designated range. Replicate data-dependent runs frequently re-sample the most abundant peptides in the mixture, so in restricting precursor selection to specified *m/z* ranges, the binning and tiling methods are able to use this time to sample low-intensity precursors, gaining sampling depth and maximizing the number quantifiable identifications (**Suppl. Figure 4.2**). **Figure 4.2 A** compares the peak depth of precursors selected for MS^2 analysis throughout one of the DDA control runs and throughout one of the tiling runs (5-segment tiling analysis, i.e. 5-tiling analysis) performed in the experiments above. Peak depth is a measure of sampling depth; if all of the species observable in an MS^1 spectrum are ranked in decreasing order of spectral intensity, the rank of a precursor selected for MS^2 analysis is the peak depth of that precursor (e.g., the most abundant ion has a peak depth of 1). While the DDA control barely sampled precursors above a peak depth of 250, the 5-tiling method sampled precursors to a peak depth of nearly 2,000 (**Figure 4.2 A**). Both the tiling and the DDA methods identified the most abundant yeast peptides, and the boost in unique peptide identifications achievable with the 5-tiling method was the direct result of the method's ability to evaluate low-level peptides that were not even considered by the DDA control method. This is displayed in **Figure 4.2 B**, which shows that the discrepancy in identifications between the two methods was the greatest for low abundance yeast proteins.

Label-free Quantification with Tiling vs. Data-dependent Methods

Our 5-tiling method consistently outperforms normal DDA in the identification of unique peptides and proteins from an unfractionated sample; however, it fails to surpass the number of identifications gained through sample fractionation. We sought to directly compare the number

of quantifiable peptides and proteins achievable using off-line fractionation and using our novel 5-tiling method. Yeast cell lysate was digested with trypsin and divided in two: one half was left unfractionated, while the other half was subjected to reverse-phase fractionation to generate five samples containing approximately equal amounts of material. The unfractionated sample was analyzed using five replicate DDA runs and the 5-tiling method, while each of the fractionated samples was analyzed once using DDA. The results are illustrated in **Figure 4.3**. Once again, the 5-tiling method produced more peptide and protein identifications than were produced by replicate DDA runs (36% and 56%, respectively), but fractionation outperformed the 5-tiling method by 37% and 34%, respectively.

Just as the 5-tiling method allows for the identification of low-level peptides absent from DDA analyses, fractionation reduces sample complexity to enable the selection and identification of precursors that remain unselected in the complex MS^1 spectra of the unfractionated sample (**Figure 4.3 B**). Sample fractionation is, therefore, the best strategy for achieving maximum protein identifications, as long as sufficient time and resources are available. Aside from the challenges associated with the reconstruction of chromatograms over multiple fractions, the major drawback of conducting MS^1-based quantification with fractionated samples is the amount of analysis time required for the acquisition of replicate analyses for reproducibility. Our 5-tiling method analyzes the same unfractionated sample in all five experiments. Since full MS^1 scans are obtained throughout each analysis in the 5-tiling method, a peptide identified in one experiment can, theoretically, be quantified in all five experiments. To achieve the same quality of quantitation (number of replicates) as the 5-tiling analysis of our unfractionated sample, the DDA analysis of our fractionated sample would require 25 (5×5) experiments. Furthermore, this elevated analysis time requirement grows considerably as the number of fractions obtained, or the number of samples being compared, increases.

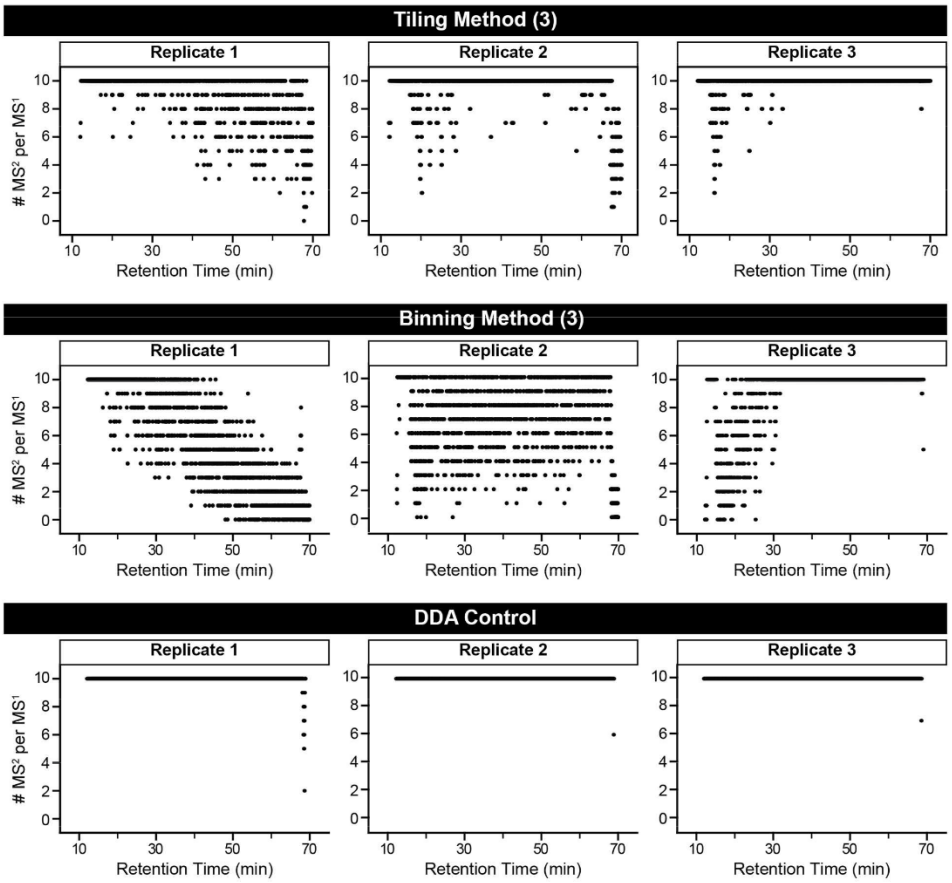
Figure 4.3 A compares the number of quantifiable peptides and proteins identified in the DDA, 5-tiling, and fractionation experiments, performed in the above investigation. Additionally, each plot depicts the number of replicate experiments in which each peptide or protein group could be quantified. A peptide was only considered quantifiable if it had at least one observed isotopic distribution in an MS^1 scan that matched its theoretical isotopic distribution with a ProteoIQ quant score of 0.5 or higher. Although the increase in peptides identified by the 5-tiling method came from lower-level proteins, the MS^1 spectral quality of these less abundant species was only marginally sacrificed; 94.8% (8,342 of 8,795) of the peptides identified by the 5-tiling method were quantifiable, compared to the 98.5% (6,353 of 6,448) quantification rate of the DDA, and the 97.0% (11,666 of 12,032) quantification rate of the fractionated analyses.

As is illustrated in **Figure 4.3 A**, all quantitative information gleaned from the five analyses of the fractionated sample is grouped into one replicate. Any peptide seen in two or more experiments is indicative of variable peptide elution over multiple fractions, a factor which complicates post-acquisition analysis. In the same amount of time, our 5-tiling method collected quantitative information in quintuplicate for ~72% of all quantified peptides and proteins. Although the same percentage of identifications could be quantified in quintuplicate for the DDA analyses, the 5-tiling method successfully quantified 31% more unique peptides, and 52% more unique protein groups, than the replicate DDA experiments. **Figure 4.3 C** shows a characteristic example of a peptide that was not identified in any of the DDA experiments, but was identified by the 5-tiling method and quantified across the MS^1 scans of all 5-tiling runs. Overall, there were more proteins quantified in all 5-tiling replicates than were even *identified* by DDA (**Figure 4.3 A**). This establishes our 5-tiling strategy as advantageous for the maximization of peptide and protein quantification using MS^1-based label-free methods.

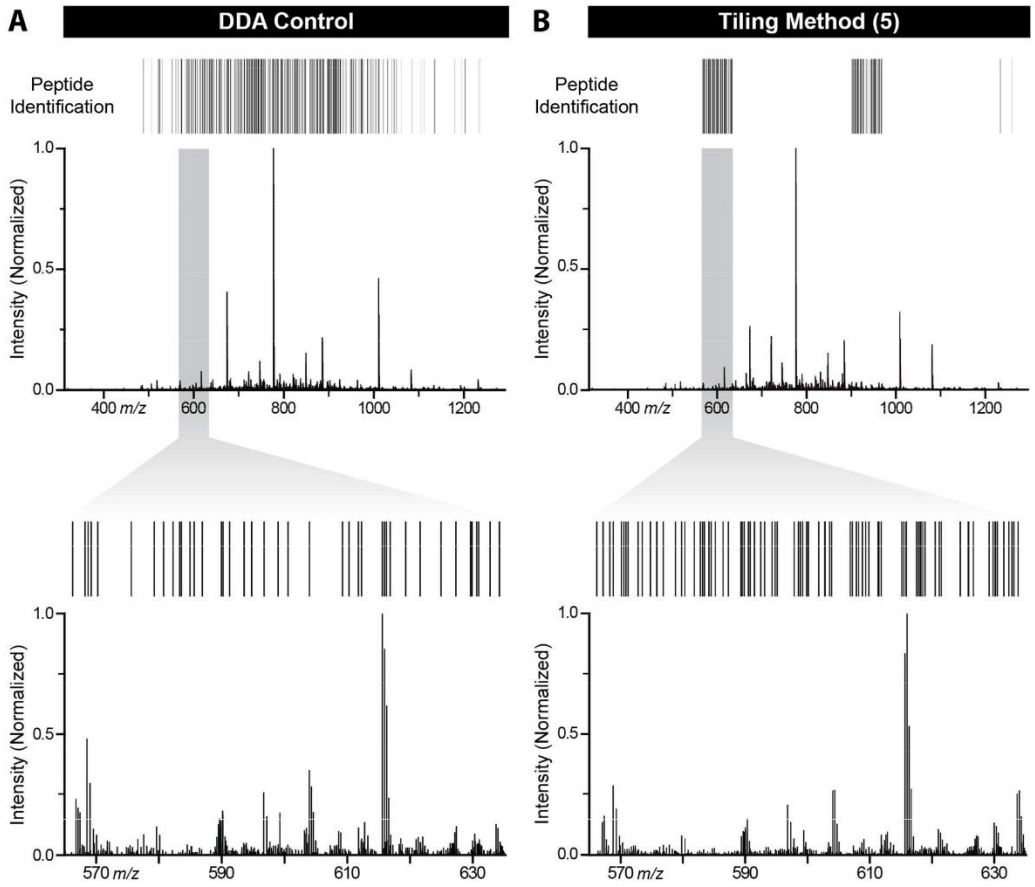
Conclusion

We present a method which restricts precursor selection to sub-segments – 'bins' or 'tiles' – of the full MS^1 mass range. Each *m/z* segment is interrogated over separate runs, but, full *m/z* range MS^1 spectra are acquired throughout each experiment. This enables the run-to-run preservation of MS^1-based label-free quantification capabilities for all peptides identified in the analysis, while increasing sampling depth to yield more quantifiable peptide and protein identifications than can be obtained using traditional data-dependent acquisition. Unlike typical data-dependent methods, our binning and tiling strategies require multiple experiments to interrogate peptides from the full MS^1 mass range. This additional time requirement, however, does not only lead to a boost in identifications; precursor selection is restricted to sub-sections of full MS^1 mass range scans, meaning that, even though a peptide may only be interrogated in one experiment, it can be quantified in every experiment. This indicates that the vast majority of the peptides identified by our binning and tiling methods will also be quantified in technical replicate.

Online, off-line, and gas-phase fractionation techniques reduce sample and MS^1 spectral complexity to gain more peptide identifications than can be attained by typical data-dependent acquisition. These strategies, however, are not ideally suited for MS^1-based label-free quantitative analysis, as they complicate chromatographic alignment for accurate XIC comparisons, and require substantial instrument time for the acquisition of replicate data. These binning and tiling strategies are, therefore, useful to the label-free quantification community, as they provide an effective way to simultaneously boost protein identifications and acquire quantification data in technical replicate in a cost- and time-efficient manner.



Supplementary Figure 4.1. Sampling efficiency of Tiling, Binning, and normal data-dependent methods. Each data point represents the number of precursors selected from an MS1 spectrum for MS2 analysis. Data points were obtained over a 60 minute period for each experiment in the Tiling, Binning, and DDA control methods. Any data point that is less than ten indicates that, at that particular point in the experiment, maximum sampling depth has been reached.



Supplementary Figure 4.2. Tiling method samples and identifies more low level peptides than normal data-dependent acquisition. Full MS1 scans from retention time 50.96 minutes in A. replicate 5 of the DDA control runs and B. replicate 5 of the Tiling runs. Each precursor peak identified in the respective analyses is denoted by a solid black line located directly above the associated peak. The greater sampling depth of the Tiling method is reflected by the higher density of peptide identifications in the regions of interest. One such region (*m/z* 565 – *m/z* 635) has been enlarged to emphasize the discrepancy between the two methods. This highlighted region is one of the *m/z* selection ranges within the Tiling method's 5th segment. Note that at this point in the gradient, the majority of peptide identifications obtained by the Tiling method reside in the first two units of the 5th segment.

	DDA Control		Tiling Method (5)		Fractionation (5 Fractions)	
Experiment	Peptide Spectral Matches	Unique Peptides	All Peptide Spectral Matches	Unique Peptides	Peptide Spectral Matches	Unique Peptides
1	6,914	4,702	2,997	2,175	5,011	3,031
2	6,731	4,638	3,051	2,211	4,747	2,900
3	6,646	4,576	2,929	1,932	4,982	3,047
4	7,121	4,906	2,914	1,843	5,038	2,937
5	7,140	4,822	3,206	2,058	4,933	2,857
Sum	**34,552**	**23,644**	**15,097**	**10,221**	**24,711**	**14,775**
Overall	-	**6,448**	-	**8,795**	-	**12,032**

Supplementary Table 4.1. Number of peptide identifications (both peptide spectral matches and unique peptides) obtained within each of the five replicates of the normal data-dependent, Tiling, and Fractionation methods.

Acknowledgements

We thank AJ Bureta for assistance with figure illustrations, and A. E. Merrill for culturing the yeast cells. This work was supported by the National Institutes of Health Grant GM080148 to J.J.C. C.E.V. was supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM T15LM007359).

Chapter 5: Design, Manufacture, and Characterization of Column Thermostats

Abstract

This chapter describes the design, production, and characterization of custom nano-HPLC column heaters for use in our improved proteomics analysis pipeline. We improved every aspect of our LC-MS based proteomics assay, including the chromatographic separation. We found smaller packing material and longer columns to be very beneficial but these improvements in separation efficiency cause an increase in pressure that is beyond the capability of most HPLC systems. Increasing the temperature at which the separation is carried out reduces the backpressure due to a decrease in solvent viscosity with increasing temperature. Therefore, a need for robust, affordable, and easy to use column ovens arose. We found that commercial offerings did not fit our needs exactly or required considerable modification of our mass spectrometer ionization sources. Consequently, I developed custom column thermostats, a process that is described in this chapter. The column heaters have been in use for over two years, have been improved several times, and have become an integral part of our lab's proteomics workflow.

Introduction

Shotgun proteomics by LC-MS has improved greatly over the past several years and it is now possible to identify several thousand peptides in one analysis taking under two hours. [54,97,102,103] Much of this gain in proteome coverage has been driven by advances in mass spectrometer instrumentation.[12,104-106] However, complete proteome coverage as opposed to complete genome sequencing remains elusive for several reasons among them the great dynamic range in protein expression, alternative splicing, and post translational modifications.[107] Improved mass spectrometers are also not the only path to increased proteome coverage – every aspect

of the workflow must be optimized. This includes sample preparation, chromatographic separation, and operating parameters of the mass spectrometer. In 2013 we embarked on a quest to improve our overall proteome coverage achievable in two hours of total instrument time by improving each part of the process and thus benefitting from the aggregation of marginal, or sometimes significant, gains (**Figure 5.1**). One of the areas we set out to improve was chromatography. We found that significant benefit, measured as increased numbers of identified peptides per unit time, can be derived from using smaller diameter packing material and longer columns. However both of these changes increase the column backpressure significantly. Backpressure varies with the square of the particle diameter and linearly with column length, according to **Equation 5.1**.[108]

$$P \propto \frac{\eta * F * L}{r^2 * d_p^2} \; eq. 5.1$$

Here, η is solvent viscosity, F is flow rate of the mobile phase, r is column radius, L is column length, and d_p is particle diameter. However, the reward for making these changes is also significant. Column plate number, N increases inversely with particle diameter d_p and is directly related to column length column length according to **Equation 5.2** below.[2]

$$N \propto \frac{L}{d_p} \; eq. 5.2$$

Thus the achievable minimum peak width of a chromatographic column for any given compound is proportional to the square roots of its length and packing material diameter as shown in **Equation 5.3.**

$$W \propto \sqrt{N} = \frac{\sqrt{L}}{\sqrt{d_p}} \; eq \; 5.3$$

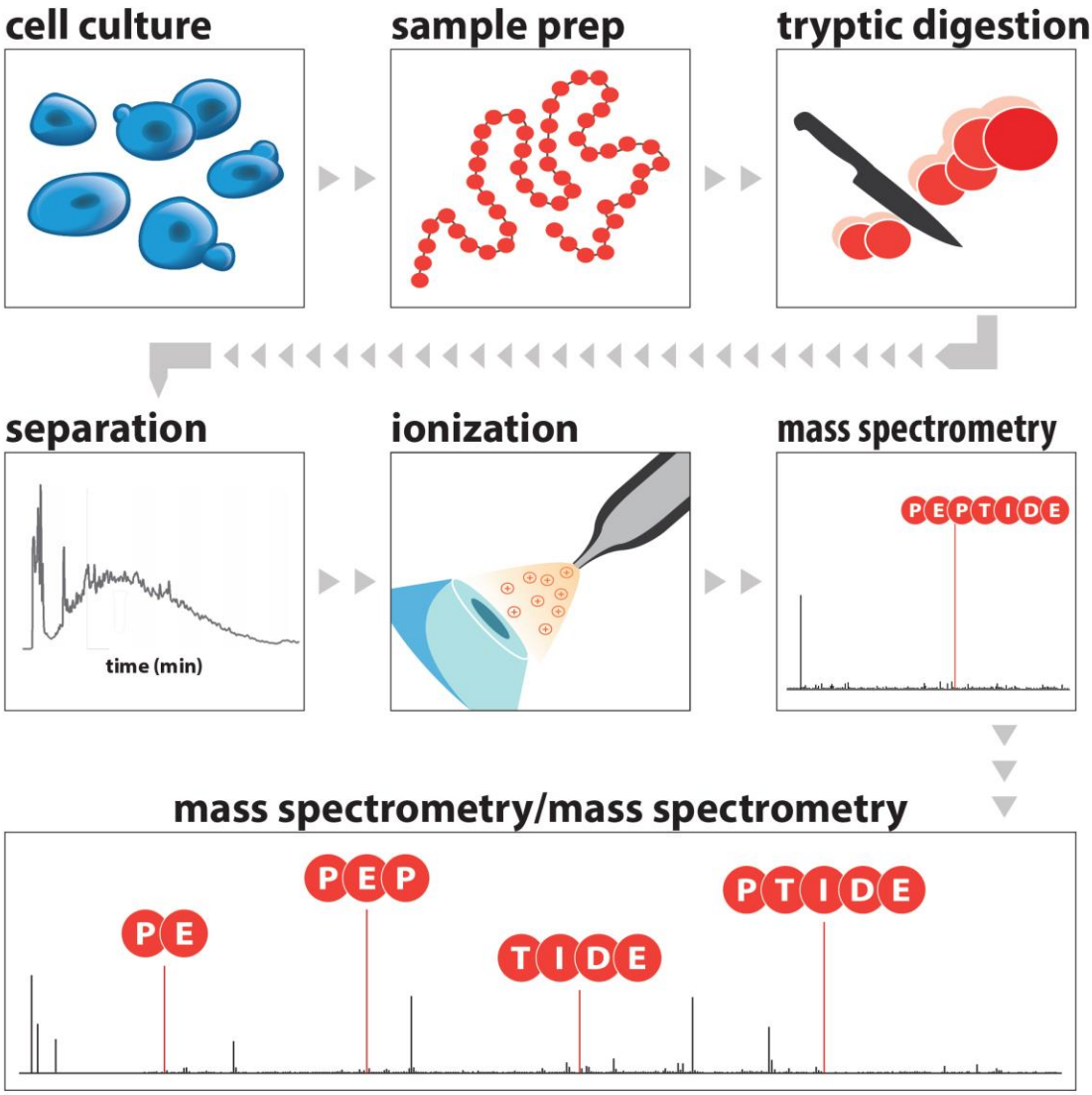


Figure 5.1 The proteomics pipeline in the Coon Research Group. Protein is extracted from cells or tissue and is then digested, usually using trypsin. Then peptides are separated by UHPLC and the column effluent is introduced into the mass spectrometer by positive ion mode ESI. Full scan mass spectra are acquired giving accurate m/z values and precursors to select for fragmentation. Tandem mass spectrometry provides the ions used for sequencing of peptides.

Having narrower peaks causes fewer peptide elution profiles to overlap, i.e. fewer peptides elute at the same time. Presenting the mass spectrometer with fewer precursors at the same time allows less abundant peptides to be sampled and reduces isobaric interferences and ion suppression.[109] **Figure 5.2** shows empirical data on the effect of packing material on measured backpressure while holding all other variables constant. The backpressure doubled when the particle diameter decreased from 5 µm to 3 µm, which is less than the predicted change of a 2.8 fold increase. This is due to the 3 µm material used having larger pores. The backpressure doubled yet again when a 2.7 µm core-shell particle was used, which is roughly equivalent to using the sub 2 µm fully porous particle, which we eventually selected (Waters BEH, 1.7 µm, C18). This material was not available at the time of data collection. It became necessary to attenuate the increase in backpressure, since many HPLC systems have pressure limits around 10,000 psi.

The variable that can be manipulated to reduce backpressure back down to manageable levels is solvent viscosity, η. Viscosity can be reduced by increasing the temperature at which the separation takes place. **Figure 5.2 B** shows data collected at varying flowrates and temperatures. A reduction in pressure of approximately 38 % can be achieved by modestly raising the analysis temperature from ambient to 50˚C. To achieve this, column thermostats for the capillary columns were needed that can easily interface with the two different types of nano-ESI ionization sources used in our research groups and that can accommodate columns of varying lengths. The types of columns used in our laboratory are 15 – 50 cm long polyimide-coated fused silica capillary columns with an outer diameter of 360 µm and an inner diameter of 75 µm. Commercial heaters were available but they either didn't fulfill some of our requirements or required electrospray source modifications to interface with our mass spectrometers in addition to being very expensive. We decided to develop our own devices and I took on the task of designing, manufacturing, and testing two types of devices to be used in our research group.

Here, I describe the design, fabrication, and performance of one of the types of column heater that is now in routine use on our mass spectrometers outfitted with NextGen ESI sources (Thermo Fisher Scientific, San Jose, CA).

Methods

This section will provide an account of the design process. The considerations that influenced the specific choice of control system, heating elements, safety systems, and physical dimensions of parts will be detailed. The process can be divided into three parts: (1) Designing of the metal parts using computer aided design software and rapid prototyping; selection of hardware for control, heating, and safety systems; and sourcing of parts. (2) Fabrication of the heaters using subtractive manufacturing and circuit soldering. (3) Performance characterization, safety tests, and continued monitoring.

Design

All metal and polymer parts were designing using Autodesk Inventor CAD software (Autodesk, San Rafael, CA). The detailed design of the parts was informed by measurements of the space constraints imposed by the ESI source and by the principle of design for manufacturability. Parts were designed to require a low number of unit operations to manufacture and the geometry was designed to make production on a standard 3-D milling machine unproblematic. After the initial design phase a simple prototype representing the outer dimensions of the finished product was made from polyethylene and a suitable fit to the mass spectrometer was verified. The final materials were selected for thermal performance, mechanical performance, and manufacturability. The main body of the heater was made from aluminum alloy, due to the material's ease of machining and high specific heat. The chassis material is polycarbonate because of its high glass transition temperature, its low thermal conductivity compared to aluminum, its toughness, and its relative lack of creep.[110]

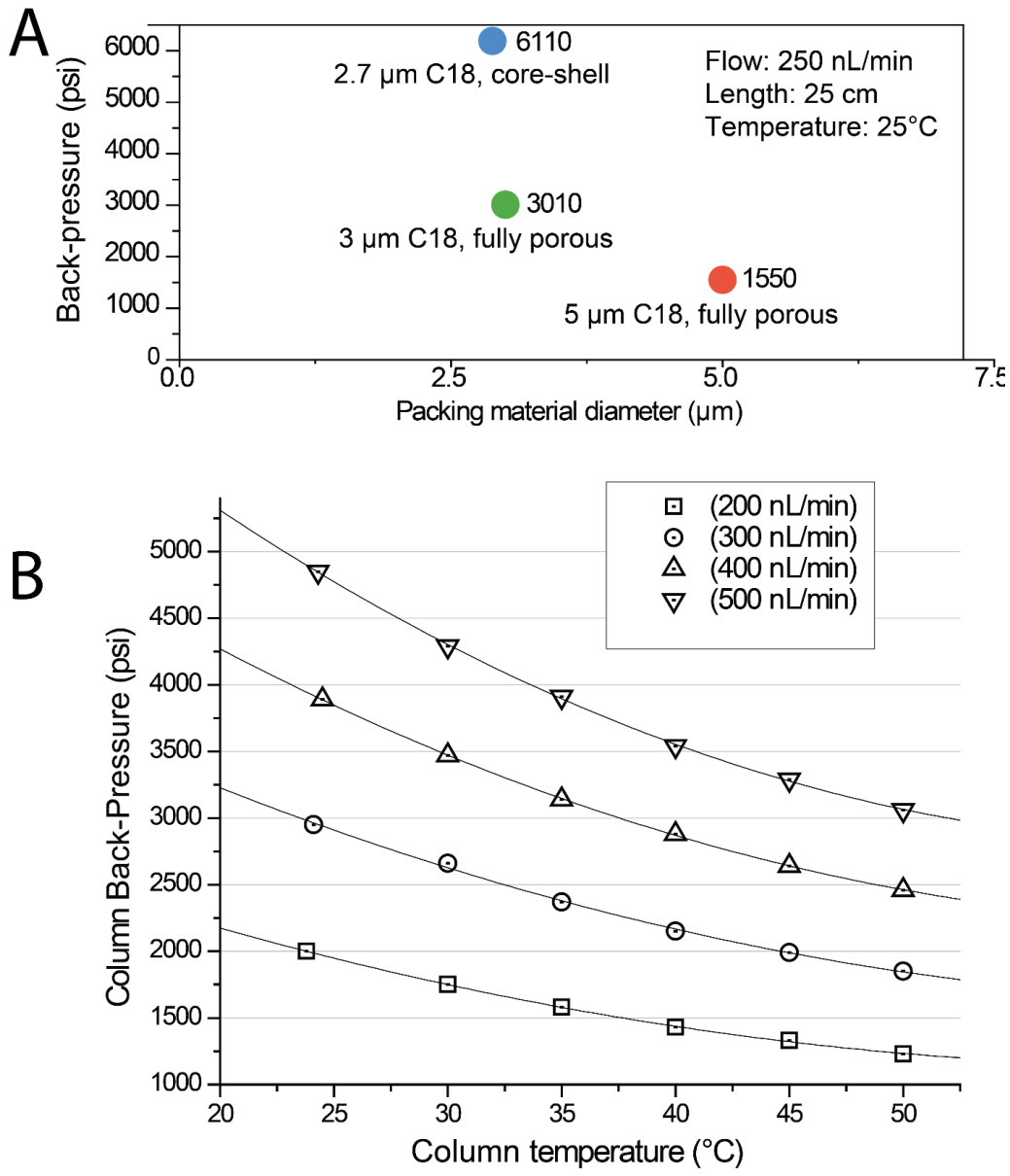


Figure 5.2 Backpressure as a function of particle diameter, mobile phase temperature, and flow rate. **A** Increase in backpressure when smaller diameter packing material is used. The pressure doubles when going from 5 µm to 3 µm material and doubles again when core-shell material is used. **B** Experimental data showing the reduction of pressure when the temperature is increased. A significant decrease in backpressure can be effected by raising the temperature, thus allowing the use of longer columns and smaller particles.

Next, the heating, control, and safety systems were designed, followed by selection of specific parts. Considerations included longevity, reliability, and safety. For example, since the heating elements have to turn on and off many times at a relatively short cycle time of several seconds, solid state relays had to be used (120D3, Opto 22, Temecula, CA) .[111] The precise maintenance of temperature is controlled by proportional integral derivative process control (PID) to ensure achieving the desired temperature in a reasonable time without too much overshoot and without oscillations of the process value around the set value (32B-23 controller, Dwyer, Michigan City, IN) .[112] A 20W resistive foil heater equipped with a pressure sensitive adhesive was affixed inside the heater body (KH-202/5-P, Omega Engineering, Stamford, CT). This heater provides even heating and its wattage is high enough to achieve and maintain the temperatures around 60˚C desirable for chromatography but low enough to be incapable of heating the oven above the glass transition temperature of the polycarbonate chassis of approximately 145˚C, making the device intrinsically safe. A 2" resistance temperature device (RTD) was employed for providing feedback to the PID controller (PR-20-2-100-1/8-E-G, Omega Engineering). A platinum RTD was chosen over a thermocouple because of the device's lower drift and higher precision compared to thermocouples.[113] The PID controller was outfitted with an RS-485 to USB interface (GC-ATC-820, gridconnect.com) to allow remote control by the mass spectrometer's instrument computer via LoveLink 3.0 software (Dwyer). The oven was grounded to prevent the danger of electric shock in case of a cable failure.

Fabrication

After construction drawings were made using the CAD software, all of the parts were machined in the university's student machine shop using a band saw, milling machine, and drill press. I received safety training and general training before conducting the work. The aluminum and polycarbonate parts are shown in **Figure 5.3.** The design used a sandwich design consisting of a polycarbonate baseplate of 0.25" thickness, three aluminum layers of 0.25" or 0.125thickness,

and an aluminum lid. The baseplate is attached to the mass spectrometer's ESI source with an adapter also machined from polycarbonate. On top of the baseplate rests a 0.125" aluminum plate, followed by a 0.25" plate containing the heater, the RTD, and the grounding wire. Then the cavity that holds the column, also 0.25" inch deep, follows. This is the most complex part. It has an undercut as well as a spring loaded column retainer to allow easy adjustment of the exact position of the column spray tip. Finally, the heater has a lid and a special tee to position the chromatographic column at the entrance of the heated cavity. The majority of the work took place on a vertical milling machine using an end mill of ½" diameter. Other cutters were used for some functions. The undercut to retain the column was cut using a dovetail cutting bit and large fly cutters were used for planing and provision of an attractive surface finish of parts. End mills were run between 3,000 rpm for small end mills and 600 rpm for large fly cutters according to instruction received by the staff and to minimize chatter. Feed rates were set to achieve a clean cut with a good finish. Generally, cuts of 0.01" - 0.04" were taken per pass. Lastly, all parts were cleaned with soap, water, and isopropanol. The aluminum parts were polished in a vibrational polisher for 5 hours followed by surface treatment with 0.1 M NaOH for 5 min to provide an attractive surface finish.[114] The heater was glued to the bottom of layer 9 (**Figure 5.3**) and holes were drilled for close sliding fits and reamed to exact size to house the RTD and the dowel pins holding the micro-tee retainer (Part 2, **Figure 5.3**). The wires from the heater, RTD, and ground connection were enclosed in expandable sheathing and connected to the PID controller and power relay. A strain relief was protected the point of entry of the wires into the thermostat body. The connections to the PID controller, relay, and USB interface were either made by screw connection or soldered and the electronics were enclosed in a housing.

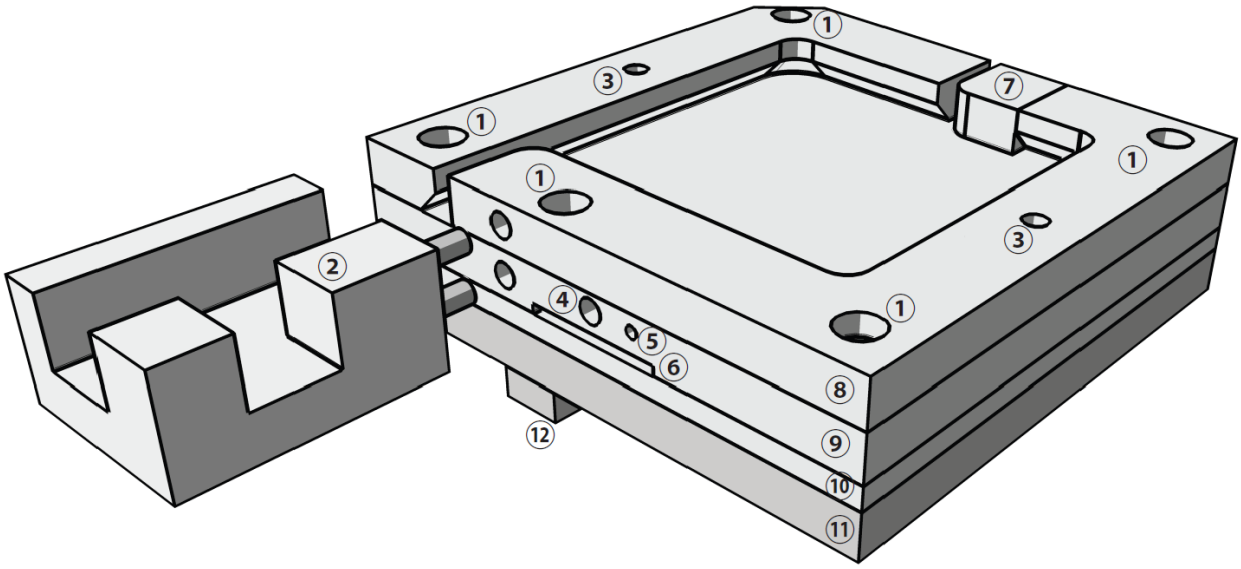


Figure 5.3 Labeled rendering of column oven without lid. ①Holes for 4-40 x ¾" screws to attach the four layers to one another ②Polycarbonate micro-tee retainer ③Holes to accept pins from lid (not shown) ④Hole for resistance thermometer ⑤Hole for grounding wire ⑥Slot for foil heater connections ⑦ Spring-loaded column tip retainer ⑧Undercut top layer ⑨Layer containing thermometer, heater (against bottom), and grounding wire ⑩Aluminum layer between heater body and polycarbonate base plate ⑪Polycarbonate base plate ⑫Polycarbonate attachment to NSI source

Performance Characterization

After final assembly, proper and safe operation of the heaters was assured by several tests. All heaters were connected to a Windows computer via USB for characterization. Remote control ability was verified by sending temperature commands from the LoveLink 3.0 (Dwyer) software and verifying that the PID controller received and acted upon those commands. The maximum allowable process value was set to 80˚C to prevent accidental setting to a temperature that would destroy chromatographic columns by an end-user. The correct type of RTD, DIN/IEC 60751, was set in the PID controller's settings. Proper heating of the oven was verified by manually checking the temperature inside the oven with a handheld thermocouple thermometer. Proper grounding of the device was verified by measuring resistance from several spots on the body to the ground prong of the oven's power cord. The PID controller was tuned using the autotune function to ensure proper ramping and stable soaking. Long-term temperature stability and proper ramping was verified by collecting temperature data and saving it to the computer (**Figure 5.4 A**). Finally, a runaway heating event was simulated by setting the set value to a very high temperature, thus causing the heating element to deliver maximum output continuously. A maximum temperature of approximately 132 ˚C was asymptotically approached over the course of 1.5 hours, which is in the safe range to have near the mass spectrometer inlet and also below the glass transition temperature of the polycarbonate chassis (**Figure 5.4 B**). The device's integrity will remain intact in such an event and the only damage that will be done is to the chromatographic column. Using a low wattage (≤25 W) heater makes for an intrinsically safe design that obviates the need for a safety system containing thermal fuses or a second RTD. This safety comes at the price of comparatively slow heating to the desired process value, e.g. it takes about 9 min to achieve a stable temperature of 60 ˚C from ambient temperature. However, the device is designed to be used to maintain a stable temperature over long periods

of time and not to quickly ramp temperatures, so a 10 min wait at the beginning of a set of analyses that may take several days was deemed negligible.

Results

The column heaters described in this chapter have been used for nearly all proteomic work in our lab since 2013. Their development played an important role in the optimization of our bottom up proteomics pipeline to enable nearly complete analysis of the yeast proteome in about one hour.[102,103] Similar devices have also been developed to be used with older models of nano-ESI sources and analytical scale columns to accommodate space constraints on a small molecule analysis setup. These models have also been used extensively for published work.[16,19,20,28] In a separate research project focusing on the improvement of each step in the proteomics pipeline (**Figure 5.1**) the column ovens allowed the use of smaller diameter packing material and longer columns. They played an important part in allowing us to increase the number of unique peptides identified from about 6000 to 21000 and the number of identified proteins from 1100 to 2600 on an Orbitrap Elite mass spectrometer using 2 hours of total instrument time (manuscript in preparation).

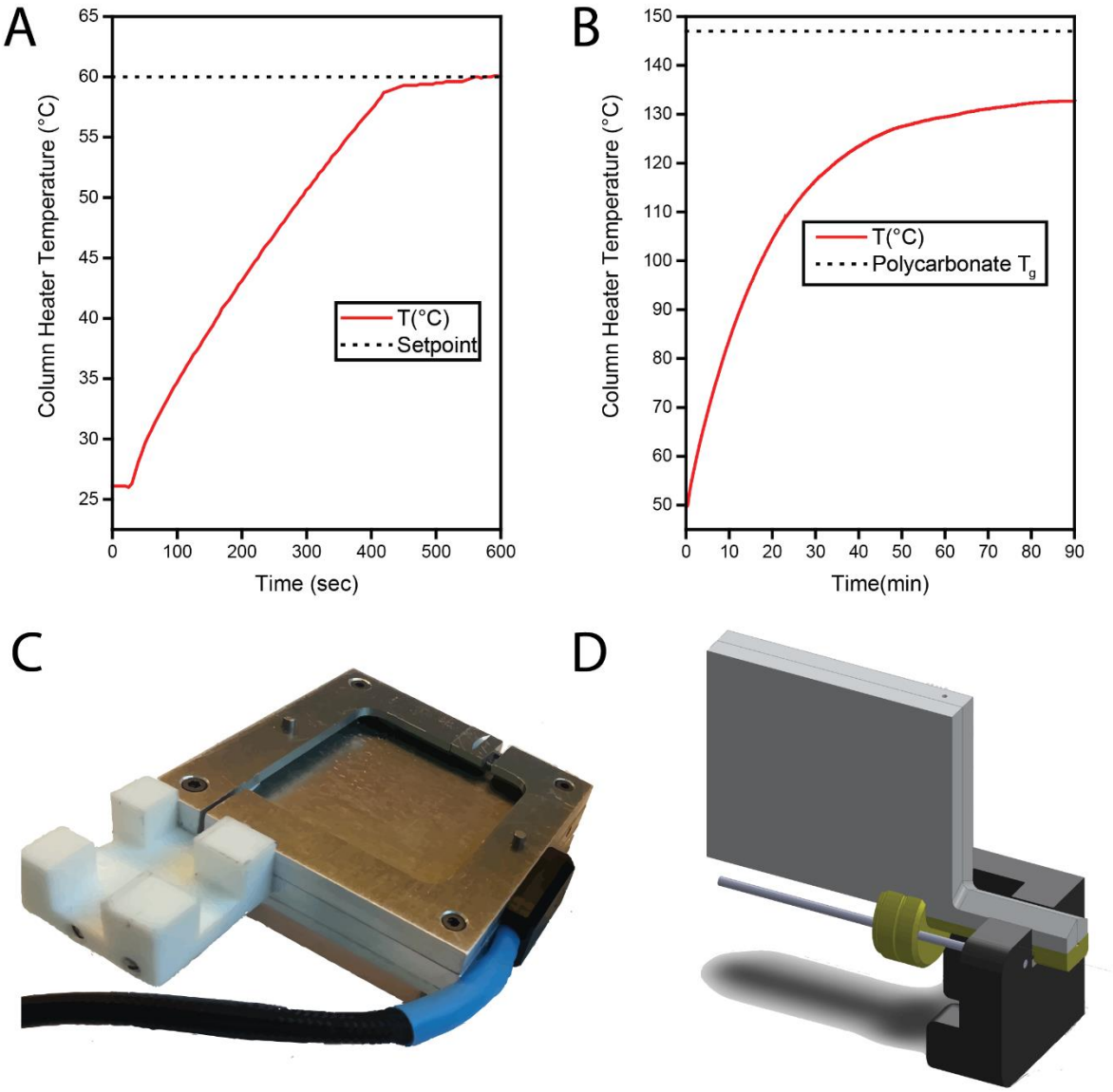


Figure 5.4 Performance characterization of the column oven and physical appearance of the two nano-HPLC models. **A** Regular operation. Heating of the device from ambient to 60˚C, a typical process value. **B** Behavior during a runaway heating event. The process value remains below the softening point of polycarbonate. **C** Photograph of the column oven (without lid) fitting our newer mass spectrometers. **D** CAD rendering of the column oven for our older mass spectrometers.

Chapter 6: Development of LC-MS Methods for Natural Substituted Aromatics

Parts of this chapter have been published:

Rahimi, A., Ulbrich, A., Coon, J. J., & Stahl, S. S. (2014). Formic-acid-induced depolymerization of oxidized lignin to aromatics. *Nature*, 515(7526), 249-252

Austin, S., Kontur, W. S., Ulbrich, A., Oshlag, J. Z., Zhang, W., Higbee, A., ... & Noguera, D. R. (2015). Metabolism of Multiple Aromatic Compounds in Corn Stover Hydrolysate by Rhodopseudomonas palustris. *Environmental science & technology*, *49*(14), 8914-8922

Keating, D. H., Zhang, Y., Ong, I. M., McIlwain, S., Morales, E. H., Grass, J. A.,Higbee, A, Ulbrich, A, ... & Landick, R. (2014). Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification. *Frontiers in microbiology*, *5*.

Lohman, D. C., Forouhar, F., Beebe, E. T., Stefely, M. S., Minogue, C. E., Ulbrich, A., ... & Pagliarini, D. J. (2014). Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis. *Proceedings of the National Academy of Sciences*, *111*(44), E4697-E4705.

Stefely, J. A., Reidenbach, A. G., Ulbrich, A., Oruganty, K., Floyd, B. J., Jochem, A., ... & Pagliarini, D. J. (2015). Mitochondrial ADCK3 employs an atypical protein kinase-like fold to enable coenzyme Q biosynthesis. *Molecular cell*, *57*(1), 83-94.

6.1: Analysis of Lignin Depolymerization Products and Fermentation Inhibitors

6.1.1 General Description of Methods

Abstract

This chapter describes LC-MS methods developed for the quantitation of lignin depolymerization products and for molecules that inhibit the fermentation of cellulosic biomass into fuel ethanol. The initial identification of these aromatic compounds by LC-MS, LC-UV, and GC-MS is also described. The analytical methods were developed for collaborations with other research groups in the chemistry department and at the Great Lakes Bioenergy Research Center. At the end of this section two papers, on which I am the second and third author, that used the methods I developed are attached.

Introduction

Lignin, an aromatic polymer formed by radical polymerization, is present at high levels in plant tissue to confer rigidity that allows growth against gravity. Almost 30% of Earth's organic carbon is contained within it making lignin a rich potential source of aromatics as an alternative to petroleum derived chemicals.[20,29,115] However, its monomers are not easily accessible and its breakdown products limit digestibility of cellulosic biomass to biofuels.[116] At present, lignin use is confined to relatively low-value areas and much of it is burned.[117] Lignin plays an important role in a biorefinery's product stream in at least two ways: Lignin has the aforementioned deleterious effect on fermentation of biomass to ethanol or other fuels but lignin could also serve as a valuable precursor for aromatics (**Figure 6.1.1.1**). Lignin could add another value stream to the biorefinery, provided economical methods for its breakdown become available. The work presented in this chapter contributed to a better understanding of lignin's role in both processes, as a precursor to inhibitory compounds and as a potentially economically valuable feedstock.

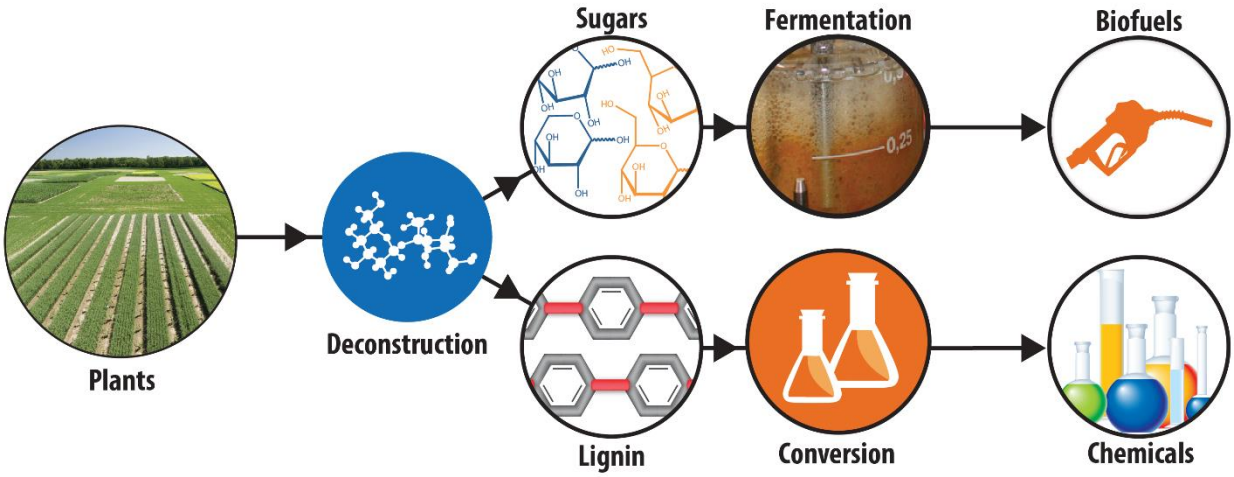


Figure 6.1.1.1 The Biorefinery Product Pipeline. Plant biomass is pre-treated to effect deconstruction and then the sugar stream for fermentation to liquid fuels and the lignin stream for conversion to value-added chemicals diverge. Residual lignin in the fermentation stream causes inhibition of fermentation. This figure is adapted from material provided to the author by the GLBRC.

Deriving transportation fuels not solely from petroleum but also from renewable resources has several benefits. Renewable fuels are more sustainable since petroleum is a finite resource, reliance on renewables reduces dependence on oil from potentially hostile nations, and because domestic innovation and job creation is spurred by renewables development. Legislation is in place in the USA to encourage a gradual weaning from petroleum-based gasoline and diesel in favor of renewable alternatives.[118,119] Already, most gasoline in the USA contains up to 10 % ethanol, mostly made from corn grain. The Renewable Fuels Standard distinguishes between *conventional biofuels* derived from starch feedstocks, such as corn or other grains, and *advanced biofuels*, i.e. fuels from cellulosic material or waste.[119] Currently, the majority of fuel ethanol is of the *conventional* type. The Energy Independence and Security Act of 2007 sets volume requirements that demand large increases in cellulosic biofuel production until 2022 but not in *conventional* bioethanol. Producing those advanced fuels from cellulosic biomass, which is not consumable, avoids the *food vs. fuel* argument that blames biofuel production for food price spikes altogether.[120,121] The stage is set for cellulosic biofuels production to increase dramatically. The Great Lakes Bioenergy Research Center (GLBRC) aims to improve the provision of cellulosic biofuels at competitive prices. However, barriers to large scale production of cellulosic biofuels remain. Alongside enzyme cost, fermentation inhibitors arising from lignin during pretreatment are a barrier to the efficient conversion of cellulose-derived sugars to ethanol or other liquid biofuels.[29,122,123] GLBRC researchers are working on developing pretreatments that minimize the occurrence of these so-called lignotoxins, on microbes more tolerant to their presence, and on detoxification strategies. In our research group, we developed the analytical methods to identify and quantify these lignotoxin fermentation inhibitors. These methods allow us to provide timely feedback on the successes or failures of efforts at the GLBRC.

The presence of lignin in the product stream does not have to be just a nuisance to be dealt with. It can also provide an opportunity for another revenue source to make production of the low-margin product fuel ethanol more economical.[124-126] Methods using depolymerization that can convert lignin into valuable industrial precursors or fuels in an economically viable fashion are required to realize lignin's potential as a sustainable and virtually limitless source of aromatics.[127] Depolymerization of lignin into a variety of precursors has been demonstrated many times before using pyrolysis, gasification, chemical oxidation, hydrolysis, hydrogenolysis, and enzymatic or biological methods.[127-129] Challenges that researchers face are: A complex mixture of reaction products whereas an industrially scalable process should yield only one or a few products; low overall conversion yields; and high costs that make the process economically unviable.[128,130] Researchers in the University of Wisconsin-Madison's chemistry department, in affiliation with the GLBRC, are working on improving lignin valorization processes and have recently developed a promising aerobic oxidation method for lignin that facilitates cleavage of lignin model compounds.[131] They then developed an efficient cleavage procedure and demonstrated it on dimeric lignin model compounds. When these methods were applied to authentic lignin help was needed to identify and quantify the products produced and we began to work on the project. We used LC-MS, GC-MS, and LC-UV to identify all major reaction products. Then we developed an LC-UV-MS based quantitation method to quickly and accurately assess product composition and lignin conversion yield. This resulted in gradual improvement of the methods to afford over 50 % conversion of lignin dry mass to a simple mixture of ten aromatic monomers.[20,132]

Methods

Measurement of Lignotoxins in Fermentation Media

The LC-MS method I developed for the quantitation of aromatic inhibitors was first published in 2014 in a paper investigating the effects of aromatics derived from ammonia-pretreated corn stover.[29] An excerpt from the paper's method section, written by Arne Ulbrich, follows below.

> The majority of phenolic compounds were determined by RP-HPLC-HRAM MS, which was carried out with a MicroAS autosampler (Thermo Scientific) equipped with a chilled sample tray and a Surveyor HPLC pump (Thermo Scientific) coupled to a Q-Exactive hybrid quadrupole/Orbitrap mass spectrometer by electrospray ionization. The analytical column was an Ascentis Express column (150 × 2.1 mm × 2.7 μm core-shell particles, Supelco, Bellefonte, PA) protected by a 5 mm C18 precolumn (Phenomenex, Torrance, CA). Mobile phase A was 10 mM formic acid adjusted to pH 3 with ammonium hydroxide and mobile phase B was methanol with 10 mM formic acid and the same volume of ammonium hydroxide as was added to mobile phase A. Compounds were separated by gradient elution. The initial composition was 95% A, which was held for 2 min after injection, then decreased to 40% A over the next 8 min, changed immediately to 5% A and held for 5 min, then changed back to 95% A for a column re-equilibration period of 7 min prior to the next injection. The flow rate was 0.3 mL/min.
>
> The HPLC separation was coupled to the mass spectrometer via a heated electrospray (HESI) source (HESI II Probe, Thermo Scientific). The operating parameters of the source were: spray voltages: +3000, −2500; capillary temperature: 300°C; sheath gas flow: 20 units; auxiliary gas flow: 5 units; HESI probe heater: 300°C. Spectra were acquired with fast polarity switching to obtain positive and negative mode ionization chromatograms in a single analysis. In each mode, a full MS^1 scan was performed by

the Orbitrap analyzer followed by a data dependent MS^2 scan of the most abundant ion in the MS^1 scan. The Q-Exactive parameters (both positive and negative modes) were: MS^1 range 85–500 Th, resolution: 17,500 (FWHM at 400 *m/z*), AGC target: 1e6, maximum ion accumulation time 100ms, S-lens level: 50. Settings for data dependent MS^2 scans were: isolation width: 1.8 Th, normalized collision energy: 50 units, resolution: 17,500, AGC target: 2e5, maximum ion accumulation time: 50 ms, underfill ratio: 1%, apex trigger: 5–12 s, isotope exclusion enabled, dynamic exclusion: 10 s.

The method described above has been in use at the GLBRC with very slight modifications for over two years and an average of 30 samples per month are analyzed by technical staff. The aromatic amides coumaroyl and feruloyl amide are the two most abundant substances that arise from AFEX pretreatment but lower concentration lignotoxins can still have very significant toxicity. Typical LC-UV chromatograms from a standard injection and from a hydrolysate broth injection are shown in **Figure 6.1.1.2 A**. The large range of concentrations of different analytes usually necessitates the injection of two dilutions of a sample for quantitation: one to achieve sufficient signal to detect and quantify lowly abundant compounds and one to reduce the signal for the amides so that it remains in the linear range of the MS detector. Typical applications of the method are quality control of synthetic hydrolysates and following the metabolism of compounds during fermentation (**Figure 6.1.1.2 B**).

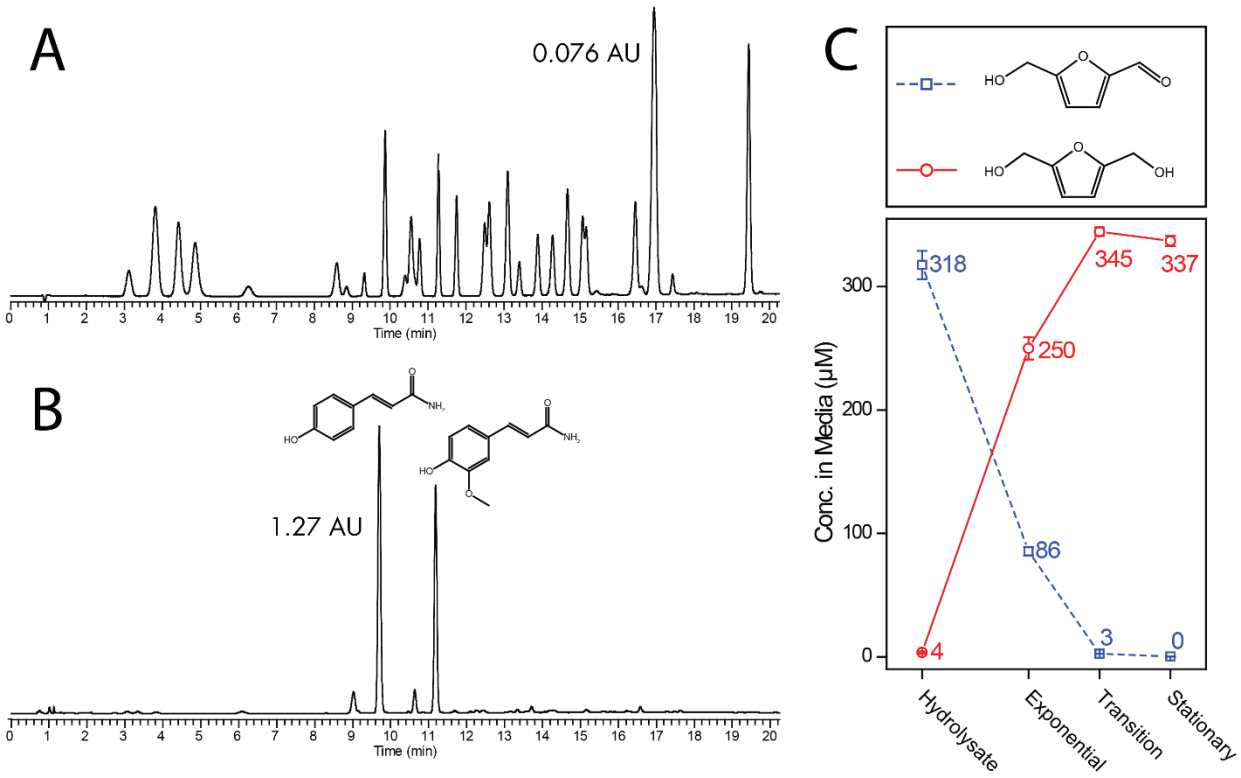


Figure 6.1.1.2 Quantitation of lignotoxin fermentation inhibitors. **A** LC-UV (for viewing clarity, actual quantitation by LC-MS) chromatogram of an injection of 25 µM combined lignotoxin standard. **B** Injection of 1:10 diluted biomass hydrolysate showing very strong peaks for coumaroyl and feruloyl amide. **C** Typical quantitation result using the LC-MS method. The reduction of 5-hydroxymethylfurfural to the diol during ethanol fermentation.

Identification of Lignin Depolymerization Products

Let us focus again on the potential benefits of having lignin in the biorefinery. After Shannon Stahl and Ali Rahimi used their method for oxidation and subsequent depolymerization on authentic lignin samples, they were unable to satisfactorily identify the products and to quantify the yields of the reactions. My first task was to establish the identity of all the major products. This was done by assembling enough evidence from several analytical techniques to make a good educated guess and then synthesizing or buying a standard to confirm the proposed identity of the unknown. The information used was retention times from GC and LC, UV/Vis spectrum, elemental composition from accurate mass, electron impact fragmentation pattern, and tandem mass spectrometry fragmentation behavior (**Figure 6.1.1.3**). The LC-UV-MS method used for identification of compounds was the same as the quantitative method described in the next section. The GC-MS conditions are described below as previously published.[20]

> The GC-MS system consisted of a Trace GC Ultra oven (Thermo Scientific) coupled to a custom-built quadrupole-Orbitrap mass spectrometer.[76] Electron impact ionization was used for all analyses. The oven was held at 60 ˚C for 3 min and then increased to 300 ˚C at 5 ˚C/min after a 1 μL splitless injection of TMS-derivatized sample diluted in chloroform. Spectra were acquired in full scan mode (50–750 Th, R = 35,000, AGC target = 1e6, max injection time = 250 ms, profile acquisition). Trimethylsilyl derivatives of all samples and standards were analyzed.

Quantitation of Lignin Depolymerization Products

In addition to identifying of the products, I developed a quantitative HPLC-UV-MS method to determine overall reaction yields and the amount of each product formed. The detailed conditions have been published and the method section as published, written by Arne Ulbrich, is below.[20]

> LC-MS quantitation of depolymerization products was performed via gradient elution from an Ascentis Express C18 column (150 mm x 2.1 mm x 2.7 µm particle size, Supelco, Bellefonte, PA) followed by integration of extracted ion chromatograms of $[M-H]^-$ ions. A quadrupole-Orbitrap hybrid mass spectrometer (Q Exactive, Thermo Scientific, San Jose, CA) equipped with a heated electrospray source (HESI II Probe, Thermo Scientific) and a photodiode array UV detector (Accela PDA Detector, Thermo Scientific) were used to detect analytes. The system was fitted with an autosampler that kept samples at 4 ˚C prior to injection (Thermo PAL, Thermo Scientific). Mobile phase A consisted of water with 0.1% formic acid adjusted to pH 3 with ammonia and B was acetonitrile. The gradient program at 0.4 mL/min flow rate and ambient temperature was: 10% B for 1 min, to 10% B over 14 min, to 70% B over 5 min, and held there for 5 min before immediately returning to initial conditions and re-equilibrating for 4 min. The mass spectrometer was operated in fast polarity switching mode, acquiring full scan (85 – 500 Th, R = 17,500, AGC Target = 5e5, max. injection times = 50 ms (+), 75 ms (-), centroid acquisition, S-lens = 50, and inlet = 320 ˚C) mass spectra in positive and negative ion mode during the same chromatographic run. The source conditions were as follows: Source heater at 350 ˚C and 60 and 15 units each of sheath and auxiliary gas. The UV detector acquired spectra (190–500 nm) at 5 Hz in addition to absorbance traces at 254, 280, and 322 nm. Filter bandwidth and wavelength step were 1 nm each for acquisition

of spectra and filter bandwidth was 9 nm for single wavelength UV trace acquisition. Filter rise time was set to 0.2 s. Samples were diluted to 1 mg/mL starting material equivalent before injection onto the LC-system and linear calibration curves were produced from 0.1 to 50 µM for all compounds. Post-acquisition data processing used the Xcalibur software suite (Thermo Scientific). All samples were normalized to vanillin-$^{13}C_6$ as the internal standard.

Results

The methods described above were used for several research projects that have either been published or have manuscripts in preparation. The full text of two of these publications is provided in the next two subsections of this thesis. The paper by Austin et al (**Section 6.1.2**) contains several figures based on the results of the lignotoxin quantitation method that represent the nature of typical results well. Results from quantitation of lignin depolymerization products can be found in the paper prepared with Rahimi, Coon, and Stahl attached in **Section 6.1.3**.

The identification of the unknown 3,5-dinitroguiaiacol will serve as an example of using the workflow described in the methods section above. First the elemental composition was determined to be $C_7H_7N_2O_6$ from a high mass accuracy negative ion mode mass spectrum. It was also noted that no positive ion mode spectrum could be obtained, excluding amine functionalities as the source of the nitrogen. The UV spectrum showed strong absorbance beyond 300 nm suggesting a nitro compound. Together with some knowledge about what types of compounds to expect from lignin this evidence led me to propose either 4,6- or 3,5-dinitroguaiacol as the identity of the product. The compounds were synthesized and then spectra and retention times of the unknown were compared to the synthesized standards. It became clear quickly that 3,5-dinitroguaiacol was the identity of the unknown. Producing this

product was not expected since lignin itself does not contain any nitrogen atoms. However, it had been shown over 100 years before our work that vanillin, which is structurally similar to lignin, can be nitrated with dilute nitric acid via a nitrovanillin intermediate to give 3,5-dinitroguaiacol.[133] A similar process of product identification was followed for all other major products (**Figure 6.1.1.3**) to arrive at a fully characterized product mixture (**Figure 6.1.1.4).** Figure 6.1.1.3 was previously published and was prepared by Arne Ulbrich.[20]

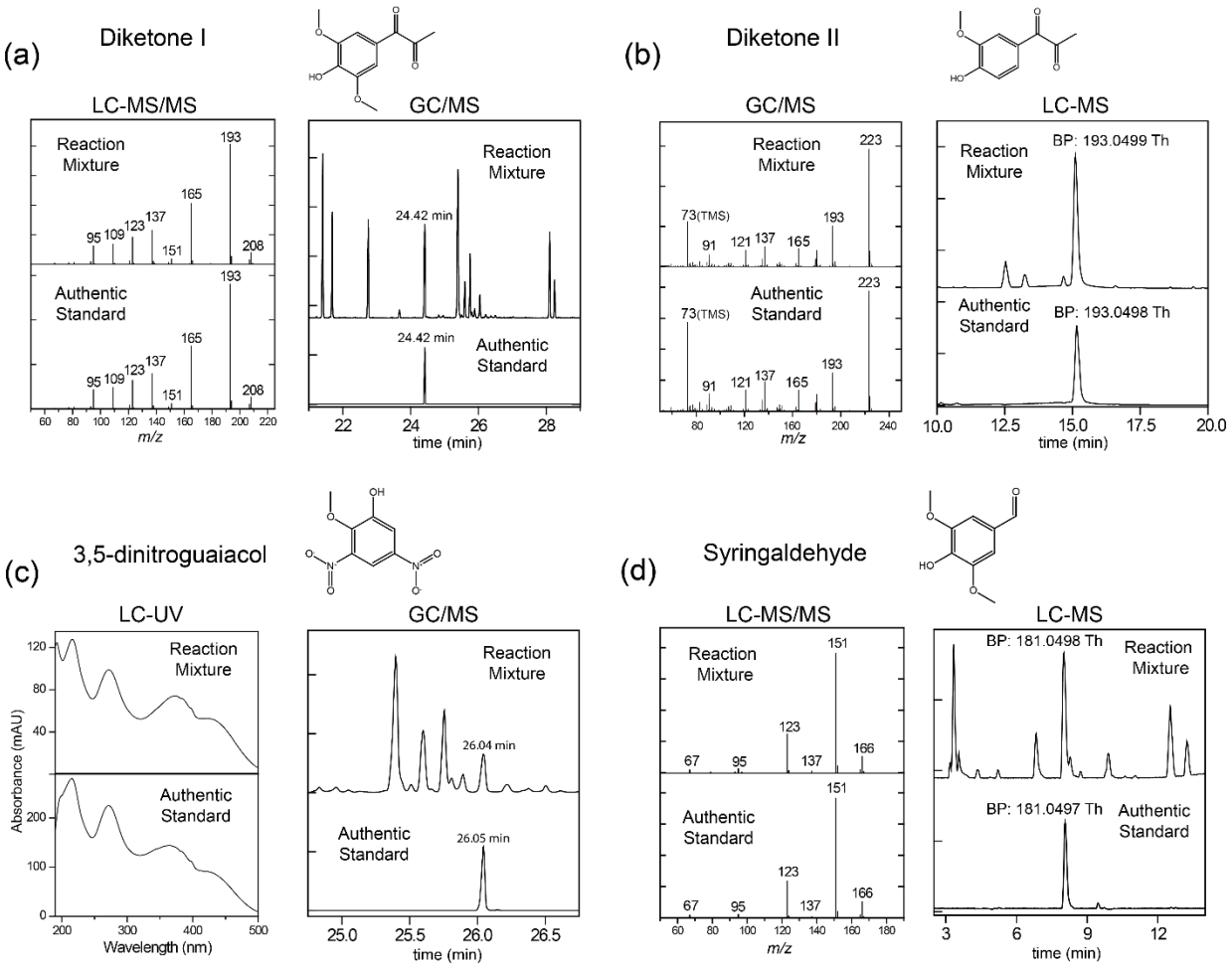


Figure 6.1.1.3 Representative samples of data used to establish the identity of lignin depolymerization products. **A** Beam-type CID mass spectra and GC-FTMS chromatograms for Diketone I. **B** Electron impact spectra and LC-MS chromatograms with exact mass for Diketone II. **C** UV spectra and GC-FTMS chromatograms for 3,5-dinitroguaiacol. **D** Beam-type CID mass spectra and LC-MS chromatograms with exact mass for syringaldehyde.

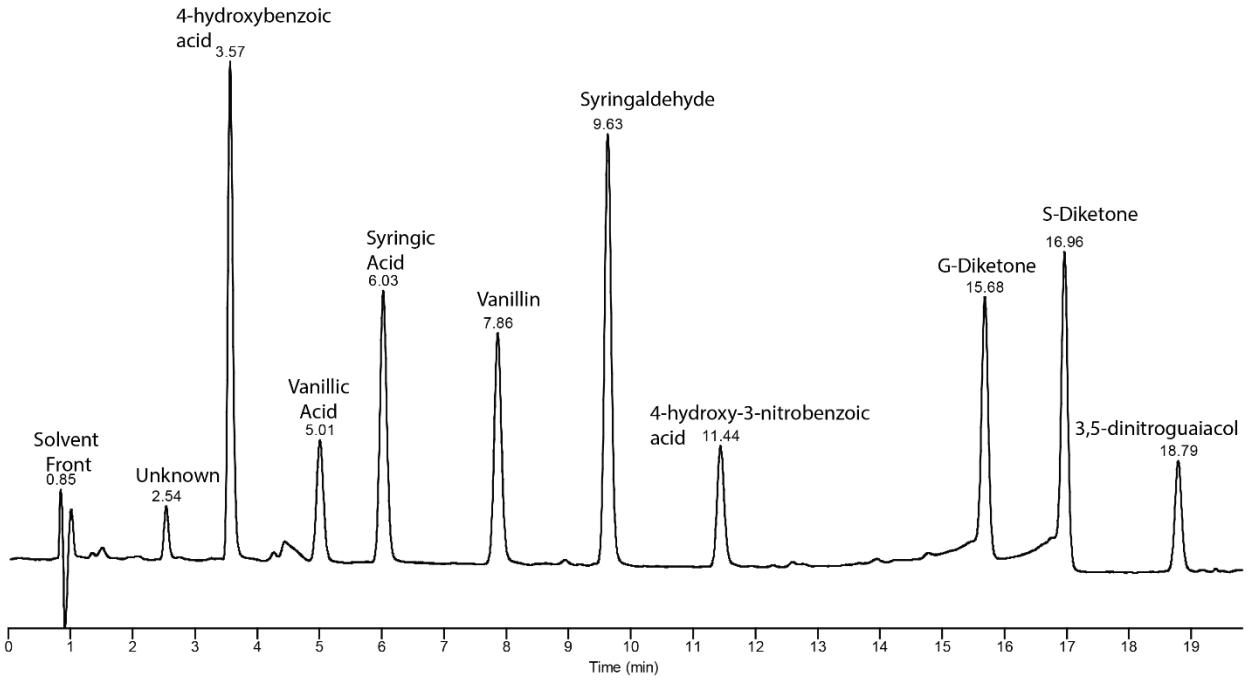


Figure 6.1.1.4 LC-UV chromatogram of a characterized sample of depolymerization products showing absorbance from 235-380 nm showing a typical mixture of products.

6.1.2 Metabolism of Multiple Aromatic Compounds in Corn Stover Hydrolysate by *Rhodopseudomonas palustris*

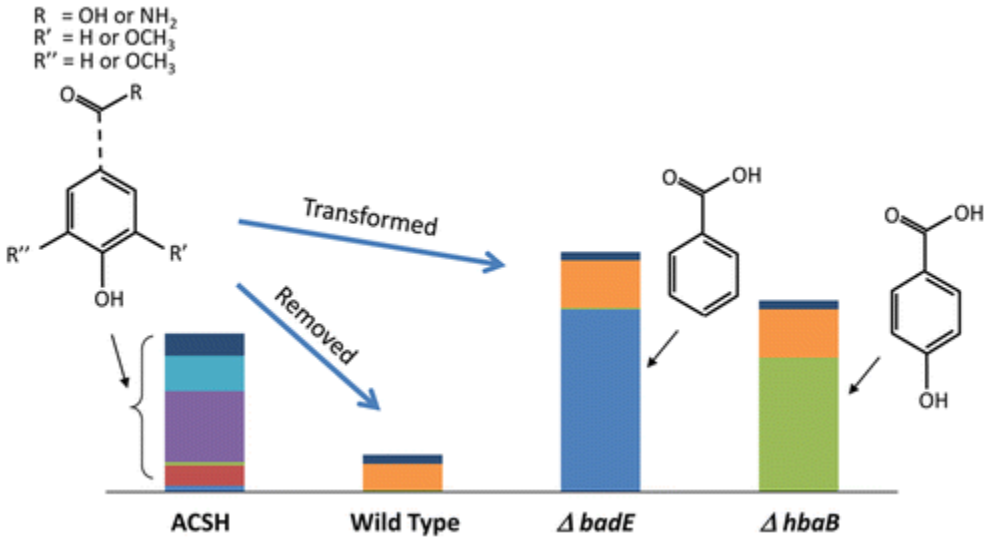
A.U. developed analytical methods, analyzed data, and produced graphics

This subsection has been published:

Austin, S., Kontur, W. S., Ulbrich, A., Oshlag, J. Z., Zhang, W., Higbee, A., ... & Noguera, D. R. (2015). Metabolism of Multiple Aromatic Compounds in Corn Stover Hydrolysate by Rhodopseudomonas palustris. *Environmental science & technology*, *49*(14), 8914-8922.

Abstract

Lignocellulosic biomass hydrolysates hold great potential as a feedstock for microbial biofuel production, due to their high concentration of fermentable sugars. Present at lower concentrations are a suite of aromatic compounds that can inhibit fermentation by biofuel-producing microbes. We have developed a microbial-mediated strategy for removing these aromatic compounds, using the purple nonsulfur bacterium *Rhodopseudomonas palustris*. When grown photoheterotrophically in an anaerobic environment, *R. palustris* removes most of the aromatics from ammonia fiber expansion (AFEX) treated corn stover hydrolysate (ACSH), while leaving the sugars mostly intact. We show that *R. palustris* can metabolize a host of aromatic substrates in ACSH that have either been previously described as unable to support growth, such as methoxylated aromatics, and those that have not yet been tested, such as aromatic amides. Removing the aromatics from ACSH with *R. palustris*, allowed growth of a second microbe that could not grow in the untreated ACSH. By using defined mutants, we show that most of these aromatic compounds are metabolized by the benzoyl-CoA pathway. We also show that loss of enzymes in the benzoyl-CoA pathway prevents total degradation of the aromatics in the hydrolysate, and instead allows for biological transformation of this suite of aromatics into selected aromatic compounds potentially recoverable as an additional



bioproduct.

Introduction

The increasing worldwide demand for energy is accelerating fossil fuel consumption, depleting natural resources, and contributing to climate change.[134] With roughly 80% of the world's primary energy supply derived from fossil fuels, there is significant interest in increasing the contribution of renewable fuels to the overall energy production portfolio. Liquid fuels generated from lignocellulosic biomass are of particular interest as transportation fuels for long-term environmental and economic sustainability.

The Energy Independence and Security Act created a roadmap for increased industrial production of biofuels from lignocellulosic biomass in the United States.[119] According to the roadmap, the production of renewable fuels from lignocellulosic biomass was expected to reach 1.75 billion gallons by 2014.[119] The actual production was only 683 643 gallons,[135] and the first generation of commercial-scale biorefineries in the U.S., to be in full operation in 2015, will not exceed an annual capacity of 50 million gallons.[136]

Clearly, major bottlenecks still exist for the cost-effective production of biofuels from lignocellulosic biomass. Some of the challenges are economic and brought about by the large amounts of fossil fuels that can now be tapped with horizontal drilling and hydraulic fracturing, which contribute to instability in the price of fossil fuels. Other challenges are technical, requiring new scientific and engineering innovation to bring transformational changes and cost reductions to the lignocellulosic biofuels industry.

One of the persistent challenges to implement cost-effective fermentation processes is the presence of plant-derived aromatic compounds and other small bioactive molecules in hydrolysates derived from lignocellulosic biomass.[137,138] Some of these molecules have been shown to diminish biofuel production by inhibiting growth and metabolism of sugars in

fermenting organisms. For instance, acetic acid is known to affect cellular processes, reduce ethanol yields, and decrease sugar consumption in wild type and engineered strains of *Saccharomyces cerevisiae*,[139,140] whereas the negative effects of a variety of furans or aromatic compounds on ethanologens such as *S. cerevisiae*, *Zymomonas mobiliz*, and *Escherichia coli* are well documented.[141-146]

The suite of inhibitory molecules in lignocellulosic hydrolysates is diverse and highly dependent on the biomass pretreatment used to produce the hydrolysates. However, aromatic compounds are reported to be found in hydrolysates independent of the plant species or hydrolysis method.[138] Strategies to overcome the effect of these inhibitory bioactive molecules range from physical or chemical removal,[147] to microbial and enzymatic degradation.[148,149] Although removal can be achieved by different approaches, in most cases the removal of the inhibitory compounds is accompanied by an undesirable decrease in the amount of sugars (e.g., 5 to 35%).[149]

In this study, we showed that *Rhodopseudomonas palustris*, a bacterium known to anaerobically degrade aromatic compounds and utilize short chain organic acids,[150] can remove inhibitory aromatics from corn stover hydrolysate, without consuming the sugars needed for biofuel production. In addition, genetic removal of selected enzymes in the benzoyl-CoA pathway, used for anaerobic aromatic metabolism in *R. palustris*, resulted in the biotransformation of the large variety of plant-derived aromatics into selected phenolic compounds that could be potentially recovered and used for other applications. To our knowledge, this is the first demonstration that the diversity of aromatics in hydrolysates can be biotransformed to a single aromatic, a strategy that could add value to lignocellulosic biomass biorefineries, where production of multiple products can aid the cost-effective and sustainable production of biofuels and chemicals from lignocellulosic biomass.[136] Thus, the data in this paper (1) provides the first demonstration of using a microorganism for metabolism and removal

of mixed aromatics present in hydrolysates produced by the ammonia fiber expansion (AFEX)[122] process, (2) demonstrates a process to remove these compounds that does not reduce the sugar content of the hydrolysates, and (3) demonstrates the possibility of not only removing inhibitory compounds from the hydrolysates, but also converting a chemically diverse suite of aromatics into single compounds (either benzoic acid or p-hydroxybenzoic acid), which can potentially be recovered as an additional valuable chemical coproduct in the biorefinery.

Experimental Section

Corn Stover Hydrolysate

AFEX-pretreated corn stover hydrolysate (ACSH) was prepared as described by Schwalbach et al.[151] and diluted ~1:3 with sterile deionized water to reach ~2% glucose content. The hydrolysates were filtered in series through 0.5 and 0.22 μm filters (Nalgene Disposable Bottle Top Filter, Thermo Fisher Scientific, Waltham, MA) prior to storage at 4 °C. Before inoculation, the pH of each hydrolysate batch, originally between 4.6 and 4.8, was adjusted to 7.0 using potassium hydroxide (KOH) pellets. After pH adjustment, the hydrolysates were filter-sterilized by passing through 0.22 μm filters.

Microbial Strains

R. palustris CGA009,[152] *R. palustris* CGA606 (CGA009-derived mutant lacking benzoyl-CoA reductase activity),[153] and *R. palustris* CGA506 (CGA009-derived mutant lacking 4-hydroxybenzoyl-CoA reductase activity)[154] were used in this study. In addition, *R. sphaeroides* 241EDD, an *R. sphaeroides* 2.4.1[155,156]-derived mutant with a modification to restore function of the *edd* gene (see the Supporting Information) was also used.

Minimal Media

Photosynthetic medium (PM), prepared as described in Kim et al.[157] and containing succinate as the organic substrate, was used for *R. palustris* growth before inoculation in hydrolysate.

Sistrom's minimal medium (SIS) containing succinate as the organic substrate, prepared as previously described,[158] was used for growing cultures of *R. sphaeroides* before inoculation into hydrolysate.

Experimental Conditions

Most experiments were conducted in an Applikon biofermenter (3L Autoclavable Microbial BioBundle, Applikon Biotechnology, Foster City, CA 94404) using 1000 mL of ACSH. In these experiments, the pH was controlled between 6.95 and 7.10 with 1 M H_2SO_4 and 10 M KOH, and the cultures were placed in front of continuous light generated by 10 W tungsten lamps. The temperature was kept at 28 °C, oxygen was removed by flushing with N_2 gas, and cell densities were measured using the Klett-Summerson colorimeter with a no. 66 filter (Klett MFG Co., NY). *R. sphaeroides* 241EDD and *R. palustris* CGA009 were pregrown in minimal media, and for each inoculation, 20 mL of culture was added to 1000 mL of ACSH.

Experiments with *R. palustris* CGA506 and CGA606 were conducted with ACSH (~1% glucose content) in the presence of 100 µg kanamycin mL^{-1}. The strain was initially grown in PM, and for each incubation, 150 µL of culture were added to 15 mL of hydrolysate. Glass reactor tubes were closed with rubber stoppers to ensure anaerobic conditions and placed in continuous illumination at 30 °C.

Analytical Procedures

Organic acids and sugars were analyzed by high performance liquid chromatography (HPLC) and quantified with a refractive index detector (RID-10A, Shimadzu) using a Bio-Rad Aminex HPX-87H column at 60 °C and mobile phase of 5 mM H_2SO_4 at 0.6 mL/min as described by Schwalbach et al.[151] Samples were prepared by filtering (0.22 µm) aliquots of the culture and diluting the filtrate 10-fold before injection into the HPLC. The majority of phenolic compounds were quantified by reverse phase HPLC – high resolution/accurate mass spectrometry, as

described in Keating et al.[29] Benzoic acid and 4-hydroxybenzoic acid were measured by high performance anion exchange chromatography–tandem mass spectrometry, using procedures also described in Keating et al.[29]

Results and Discussion

Metabolism of Plant-Derived Aromatics in Corn Stover Hydrolysates by *R. palustris* CGA009

An effective microbial strategy to remove aromatic inhibitors from hydrolysates will require an organism that specifically degrades these compounds without compromising the conversion of glucose and xylose to biofuel. In addition to being deficient in sugar utilization,[152] *R. palustris* CGA009 efficiently grows using short chain organic acids[159] and is also known for its ability to utilize aromatic compounds as sole carbon sources under anaerobic conditions, using the benzoyl-CoA pathway.[150,160,161] Entrance into this pathway (**Figure 6.1.3.1**) occurs through activation of either benzoate or *p*-hydroxybenzoate by ligation to coenzyme A (CoA). A subsequent stepwise reduction of the aromatic ring yields 1-ene-cyclohexanoyl-CoA, leading to ring cleavage and further transformations to metabolites that enter central metabolism.[153,162]

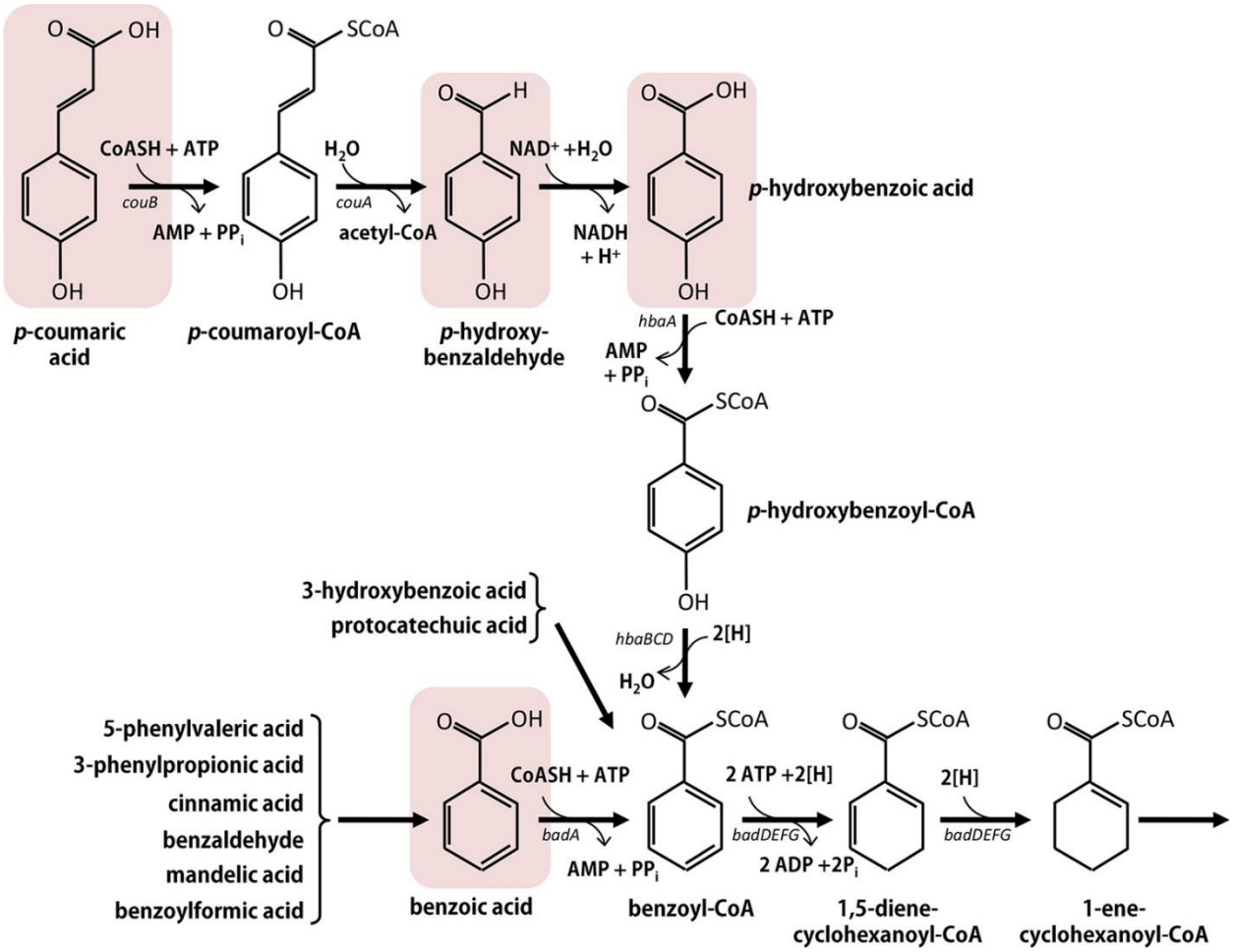


Figure 6.1.2.1. Aromatics known to be degraded by *R. palustris* CGA009 through the benzoyl-CoA pathway. The boxes indicate aromatics found in cellulosic biomass hydrolysates. References for the reactions shown in this figure are included in **Table 6.1.2.1**.

	degradation as sole carbon source		transformation in ACSH	
compound	transformed	ring fission	detected	transformed
acetosyringone			Y	N
acetovanillone			Y	N
benzaldehyde	Y(18)			
benzoic acid	Y(18)	Y(18)	Y	Y
benzoylformic acid	Y(18)	Y(18)		
caffeic acid	Y(18)	N(18)		
cinnamaldehyde	Y(18)			
cinnamic acid	Y(18)	Y(18)		
coumaroyl amide			Y	Y
p-coumaric acid	Y(18, 34)	Y(18, 34)	Y	Y
cyclohexanecarboxylic acid	Y(18)			
A-1-cyclohexenecarboxylic acid	Y(18)			
A-3-cyclohexenecarboxylic acid	Y(18)			
cyclohexanepropionic acid	Y(18)			
3,4-dihydroxybenzoic acid (protocatechuic acid)	N(18, 36)		Y	Yb
ferulic acid	Y(18)	N(18)	Y	Y
feruloyl amide			Y	Y
hydrocaffeic acid	Y(18)	N(18)		
hydrocinnamaldehyde	Y(18)			
4-hydroxyacetophenone			Y	N
4-hydroxybenzamide			Y	N
4-hydroxybenzaldehyde	Y(18)		Y	Y
3-hydroxybenzoic acid	Y(18)	Y(18)		
4-hydroxybenzoic acid	Y(18)	Y(18)	Y	Y
4-hydroxybenzoylformic acid	Y(18)	Y(18)		
DL-mandelic acid	Y(18)	Y(18)		
4-phenylbutyric acid	Y(18)	N(18)		
3-phenylpropionic acid (hydrocinnamic acid)	Y(18)	Y(18)		
5-phenylvaleric acid	Y(18)	Y(18)		
syringaldehyde			Y	Y
syringamide			Y	N
syringic acid	N(18)		Y	N
vanillamide			Y	N
vanillic acid	N(18)		Y	N
vanillin	N(18)		Y	Y

Table 6.1.2.1. Anaerobic Transformation of Aromatic Compounds by *R. palustris* CGA009[a]

[a]Harwood et al.[150] reports transformations in *R. palustris* CGA001, the parent strain of CGA009. Other compounds, not found in hydrolysates, and shown to not support anaerobic growth of *R. palustris* CGA001 are 4-aminobenzoic acid, anthranilate (2-aminobenzoic acid), catechol, 3-chlorobenzoic acid, coniferyl alcohol, 4-cresol, cyclohexanol, cyclohexanone, ethylvanillic acid, 2-fluorobenzoic acid, gallic acid (trihydroxybenzoic acid), gentisic acid, nicotinic acid, phenol, phenoxyacetic acid, 3-phenylbutyric acid, quinic acid, resorcinol, salicylic acid (2-hydroxybenzoic acid), shikimic acid, trimethoxybenzoic acid, trimethoxycinnamic acid, 3-toluic acid, 4-toluic acid, homogentisic acid, isovanillic acid, phenylacetic acid. Y = yes, N = no.

[b]Protocatechuic acid has been shown to be degraded if benzoic acid or *p*-hydroxybenzoic acid are present in the medium.[163]

Although *R. palustris* CGA009 has been shown to degrade several aromatic hydrocarbons (**Table 6.1.2.1**), little is known about its ability to utilize more complex plant-derived aromatics present in hydrolysates (**Figure 6.1.2.2**), which contain up to two methoxy functional groups and an alkyl side-chain with characteristics that depend on the methods used for biomass deconstruction.[138] **Table 6.1.2.1** summarizes the aromatics that were known to be degraded by *R. palustris* and compares them to aromatics found in ACSH. In general, *R. palustris* CGA009 has been shown to degrade completely phenolic acids without ring substitutions or with only one hydroxyl group in the *meta* or *para* position (**Table 6.1.2.1**).[150] In addition, of the phenolic acids with a propanoid side-chain, *p*-coumaric acid can be completely degraded by *R. palustris* GGA009 (**Figure 6.1.2.1**),[164] whereas only partial degradation, without ring fission, was shown to occur with aromatic acids having more than one ring substitution, such as ferulic or caffeic acid (**Table 6.1.2.1**). *R. palustris* CGA009 has also been reported not to grow on vanillin, vanillic acid, or syringic acid when present as a sole organic carbon source.[150] Growth of *R. palustris* CGA009 on the aromatic amides that are both found in hydrolysates prepared with the AFEX pretreatment[122] and known to have inhibitory effects on ethanologenic microbes[138] has not been investigated so far.

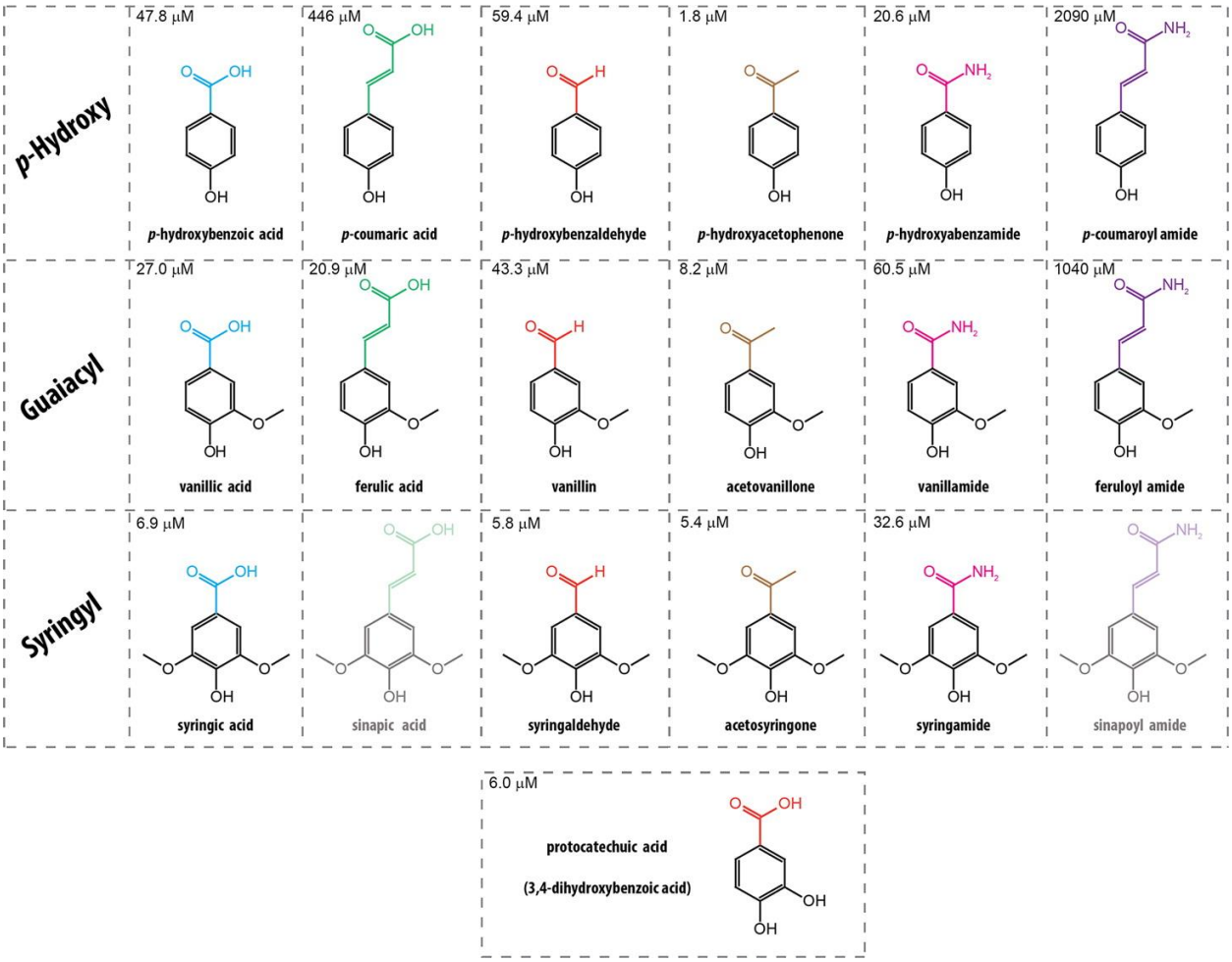


Figure 6.1.2.2. Chemical structures of aromatic compounds found in lignocellulosic hydrolysates. Concentrations in one batch of hydrolysate used in this study are indicated for each compound. The greyed out aromatics, sinapic acid and sinapoyl amide, were not detected.

Given the knowledge gap on the ability of *R. palustris* CGA009 to degrade aromatic amides and phenolics with more than one ring substitution, we evaluated the extent of aromatic transformation by this organism when grown in ACSH (**Figure 6.1.2.3**). Coumaroyl amide and feruloyl amide, the two aromatic amides present at the highest concentrations in ACSH, were removed from the medium (**Figure 6.1.2.3 A**). Concomitant with the loss of these aromatics, 4-hydroxybenzoic acid, vanillic acid, and 3,4-dihydroxybenzoic acid accumulated (**Figure 6.1.2.3 C,D**), suggesting that these aromatics are intermediates in the degradation of the propanoyl amides. Ferulic acid transiently accumulated, but was almost completely absent in the medium toward the end of the experiment (**Figure 6.1.2.3 D**). Consistent with prior knowledge on its ability to serve as a carbon source,[165] *p*-coumaric acid was removed from the medium (**Figure 6.1.2.3 C**). In addition, aromatic benzaldehydes (**Figure 6.1.2.3 E**) were also transformed, regardless of the number of methoxy groups that they contained, while the extracellular level of aromatic phenones did not change, and a small accumulation of 4-hydroxyacetophenone was observed (**Figure 6.1.2.3 F**).

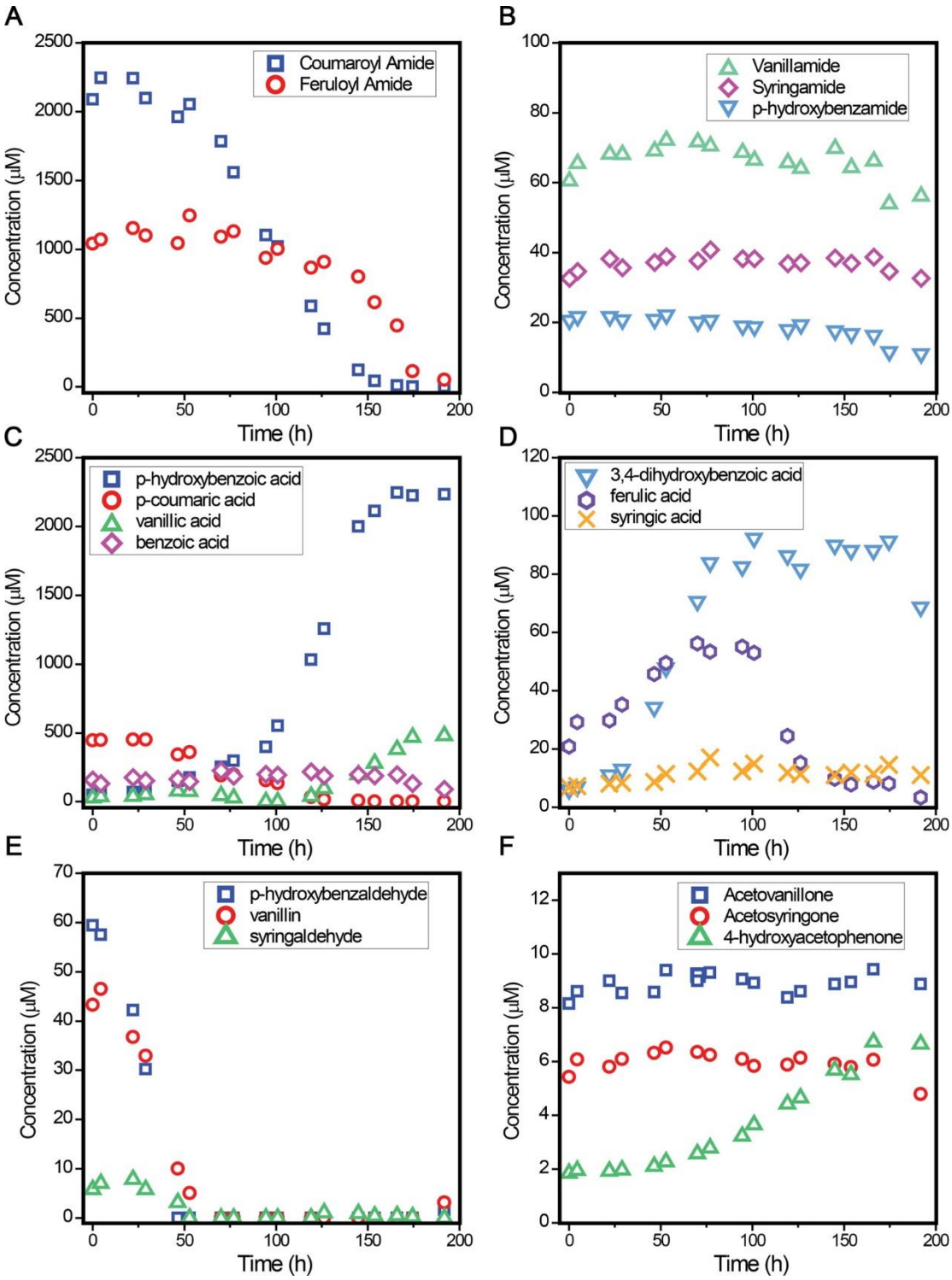


Figure 6.1.2.3. Transformation of aromatic compounds by *R. palustris* CGA009 in growth experiments using ACSH. Aromatic amides present at (A) high and (B) low concentrations; aromatic acids at (C) high and (D) low concentrations; (E) aromatic aldehydes; (F) phenones. The molecular structure of each compound is presented in **Figure 6.1.2.2**.

This experiment demonstrated for the first time that *R. palustris* CGA009 has the ability to degrade coumaroyl amide and feruloyl amide, although three other aromatic amides present in ACSH, vanillamide, syringamide, and 4-hydroxybenzamide were not transformed **Figure 6.1.2.3 B**). On the basis of the predicted pathway for *p*-coumaric acid degradation (**Figure 6.1.2.1**),[165,166] a possible route for coumaroyl amide degradation may include an initial removal of the amine group and activation to coumaroyl-CoA, followed by removal of the alkyl chain leading to *p*-hydroxybenzaldehyde and oxidation to *p*-hydroxybenzoic acid, which then enters the benzoyl-CoA pathway after CoA-ligation. The accumulation of *p*-hydroxybenzoic acid in the medium (**Figure 6.1.2.3 C**) suggests that CoA ligation of *p*-hydroxybenzoic acid is a limiting step in the use of coumaroyl amide and other aromatics in ACSH. Feruloyl amide may undergo similar transformations, with the removal of the alkyl chain after CoA ligation resulting in the formation of vanillin, and then accumulation of vanillic acid. Although it has been shown that vanillic acid is not degraded by *R. palustris* CGA001,[150] the parent strain of *R. palustris* CGA009 (CGA009 is a chloramphenicol resistant derivative of CGA001), there is evidence that other *R. palustris* strains can use vanillic acid as a sole carbon source.[150] In the experiment with ACSH, the molar concentration of vanillic acid in the medium was about one-half of the initial concentration of feruloyl amide (**Figure 6.1.2.3**), suggesting that some degradation of vanillic acid occurred. In separate experiments with more dilute ACSH and longer incubation times (see the Supporting Information), we observed complete removal of vanillic acid, and therefore, we propose that the accumulation of vanillic acid in **Figure 6.1.2.3 C** reflects a transient extracellular buildup of this putative pathway metabolite. In support of this hypothesis, we note that the extracellular accumulation of *p*-hydroxybenzoic acid is also transient (see the Supporting Information); this is not surprising because this compound is known to be metabolized via the benzoyl-CoA pathway (**Figure 6.1.2.1**).[154] Likewise, the seemingly stable

accumulation of benzoic acid in the medium over the course of this experiment is likely due to it being actively produced and consumed during the experiment, as it is also a readily degradable aromatic by *R. palustris*. The extracellular accumulation of 3,4-dihydroxybenzoic acid (protocatechuic acid) is also of note, because it is difficult to explain the source of this compound. One possibility is the removal of a methoxy group from the mono methoxylated aromatics during the degradation of vanillic acid and ferulic acid, although there is no prior knowledge that such a transformation is catalyzed by *R. palustris*. Another possibility is that protocatechuic acid is produced from the degradation of unidentified plant-derived aromatics present in ACSH. Regardless, we propose that protocatechuic acid is slowly degraded via the benzoyl-CoA pathway based on previous studies that showed that *R. palustris* CGA009 will degrade protocatechuic acid only when benzoic acid or *p*-hydroxybenzoic acid are also present, suggesting that these later compounds induce the benzoyl-CoA pathway, which does not get induced in the presence of protocatechuic acid alone.[163] The same synergy that allows *R. palustris* to degrade protocatechuic acid in the presence of other aromatics may be an explanation for the degradation of vanillic acid in ACSH.

While these experiments extend the knowledge on the range of plant-derived aromatics that *R. palustris* CGA009 can degrade, some compounds remained unutilized in ACSH. Specifically, there is no evidence that *R. palustris* CGA009 can degrade acetophenones, or dimethoxylated aromatics other than syringaldehyde present in ACSH (**Figure 6.1.2.3**).

R. palustris CGA009 Removes the Short Chain Organic Acids but Does Not Consume the Sugars Found in ACSH

The removal of inhibitory compounds from hydrolysates is needed to alleviate metabolic stress in microorganisms used for fermentative production of ethanol or other compounds from lignocellulosic biomass hydrolysates.[29,142,167] However, an effective hydrolysate pretreatment needs to selectively remove the inhibitors, while leaving the sugars and other essential nutrients

available for subsequent fermentation. As shown in **Figure 6.1.2.4**, *R. palustris* CGA009 does not consume glucose and xylose, the main sugars present in corn stover hydrolysate. In addition to metabolizing the aromatics in ACSH, acetate was the other main organic substrate that was removed from the medium in these cultures.

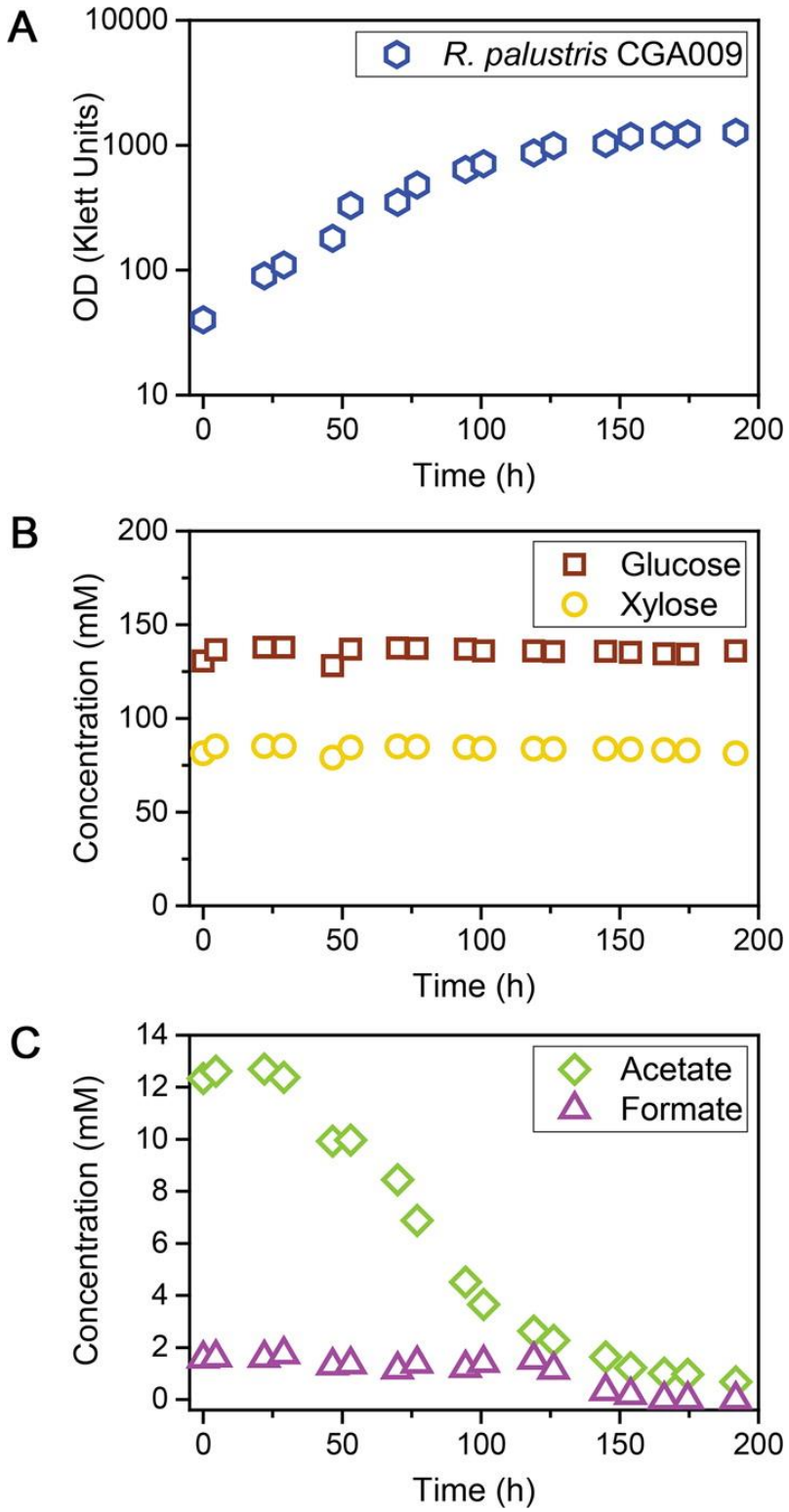


Figure 6.1.2.4. Growth (A), concentration of sugars (B), and concentration of short-chain organic acid (C) during growth of *R. palustris* CGA009 in ACSH.

Biotransformation of Aromatics with Accumulation of Benzoic Acid Derivatives Using *R. palustris* Mutants

If the most abundant aromatics present in ACSH are biotransformed via the benzoyl-CoA pathway, we hypothesized that blocking this pathway could lead to partial transformations of some aromatics, but without ring cleavage. We tested this hypothesis with *R. palustris* CGA606, a mutant with an insertion in the *badE* gene that inactivates benzoyl-CoA reductase (**Figure 6.1.2.1**) and prevents dearomatization of benzoyl-CoA.[153] Experiments with the BadE mutant showed transformation of aromatic compounds, with a significant accumulation of benzoic acid in the medium (**Figure 6.1.2.5**). Because benzoic acid is not present at high levels in ACSH and remained at low concentrations during growth using *R. palustris* CGA009 (**Figure 6.1.2.3**), we conclude that its accumulation when using *R. palustris* CGA606 is a consequence of losing BadE activity (which normally uses benzoyl-CoA as a substrate).

We also tested growth of *R. palustris* CGA506, which lacks 4-hydroxybenzoyl-CoA reductase (HbaBCD) activity,[154] in ACSH. On the basis of what is known about the benzoyl-CoA pathway, the loss of HbaBCD should block degradation of *para*-hydroxylated aromatics but not benzoic acid. In experiments using CGA506, we found loss of aromatic compounds and accumulation of 4-hydroxybenzoic acid in the medium (**Figure 6.1.2.5**). HbaBCD uses 4-hydroxybenzoyl-CoA as a substrate (**Figure 6.1.2.1**), so the accumulation of 4-hydroxybenzoic acid in the medium predicts that metabolism of the aromatics in ACSH also uses this enzyme. More importantly, these experiment show that most of the aromatics are metabolized through the common benzoyl-CoA pathway where the aromatic ring is reduced **Figure 6.1.2.1**). They also demonstrate the ability of engineered *R. palustris* strains to convert a diverse pool of aromatics into a single compound (either benzoic acid, 4-hydroxybenzoic acid, or possibly others based on the genetic block in the benzoyl-CoA pathway).

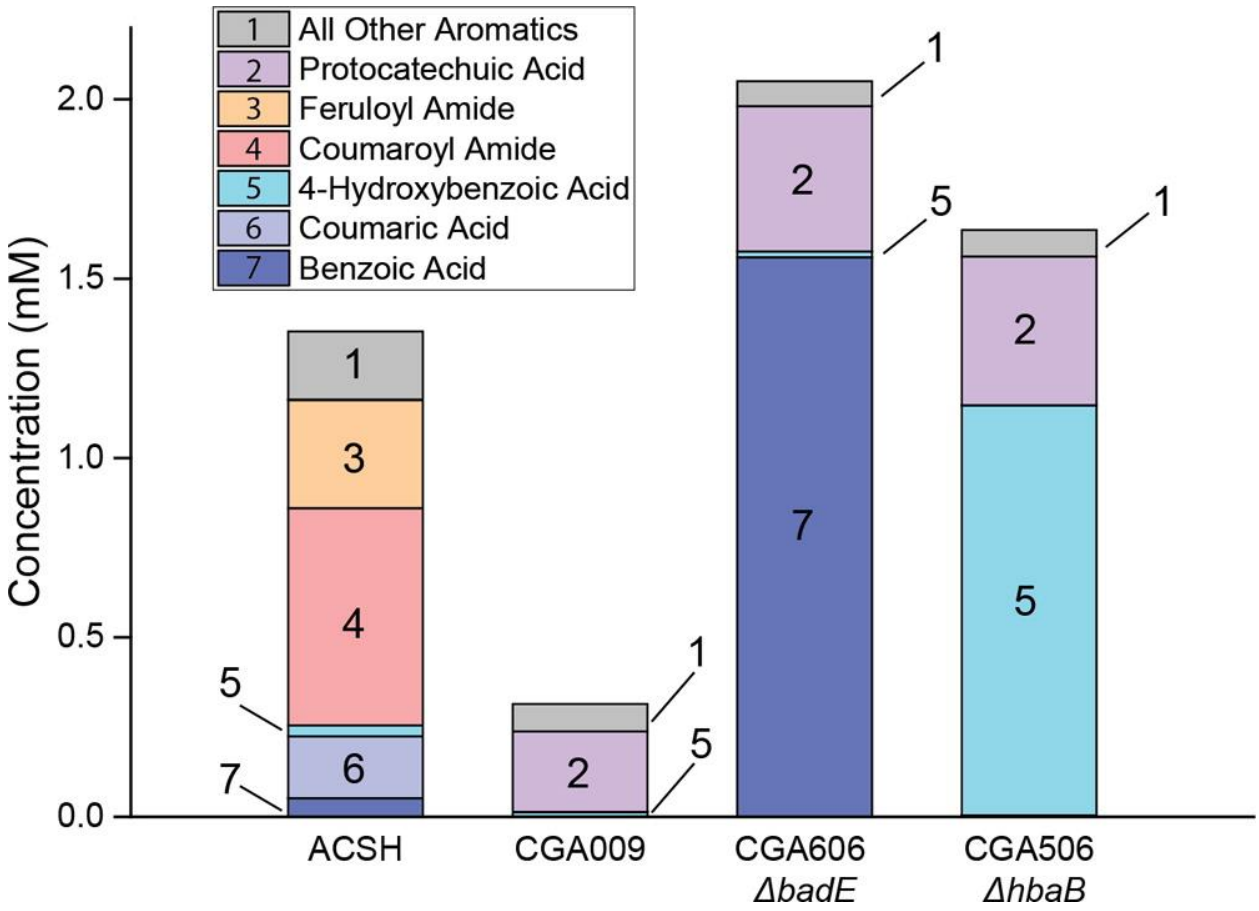


Figure 6.1.2.5. Concentrations of aromatic compounds in ACSH and extracellular concentrations of the same aromatics after growth of *R. palustris* CGA009 (wild type), CGA606 (Δ*badE*), and CGA506 (Δ*hbaB*).

Biological Removal of Aromatics from ACSH Improves Growth of a Second Bacterium

The negative effect of aromatic compounds on ethanologenic fermentations has been well documented, with *p*-coumaric acid,[168] benzoic acid,[169] *p*-hydroxybenzaldehyde,[142,144] vanillin,[144,170] 4-hydroxyacetophenone,[144] acetovanillone,[144] and aromatic amides[158] reported to be inhibitory to bacterial or yeast ethanologens. Our experiments show that most of these compounds can be removed from ACSH using *R. palustris* CGA009 (**Figure 6.1.2.3** and **Table 6.1.2.1**). Therefore, we considered the possibility that *R. palustris* can be used as a biobased method to reduce metabolic stress and improve ethanologenic or other biofuel fermentations. To test this hypothesis, we used *Rhodobacter sphaeroides*, an organism that we are investigating for production of advanced biofuels, such as long-chain fatty acids and furans.[171,172]

R. sphaeroides does not grow at the ACSH concentrations used in this study. However, when we inoculated *R. sphaeroides* 241EDD, a strain with an improved rate of glucose utilization (see the Supporting Information) in filter-sterilized ACSH that had been used to grow *R. palustris* CGA009, we found that *R. sphaeroides* could grow and metabolize the sugars. This demonstrates that biological conditioning of ACSH with *R. palustris* CGA009 allows growth and improved sugar metabolism by a second bacterium **Figure 6.1.2.6**). From control experiments where *R. sphaeroides* is grown in the presence of the aromatic compounds found in ACSH, we know that *p*-coumaric acid, *p*-hydroxybenzaldehyde, and *p*-hydroxyacetophenone inhibited glucose utilization, whereas the other aromatics were not inhibitory.[173] ACSH conditioning with *R. palustris* effectively removed *p*-coumaric acid and *p*-hydroxybenzaldehyde, but not *p*-hydroxyacetophenone, which remained at low concentrations in the hydrolysate (**Figure 6.1.2.3**). Thus, the removal of the aromatics that were present at highest concentrations in ACSH by *R. palustris* was sufficient to allow *R. sphaeroides* growth and sugar utilization.

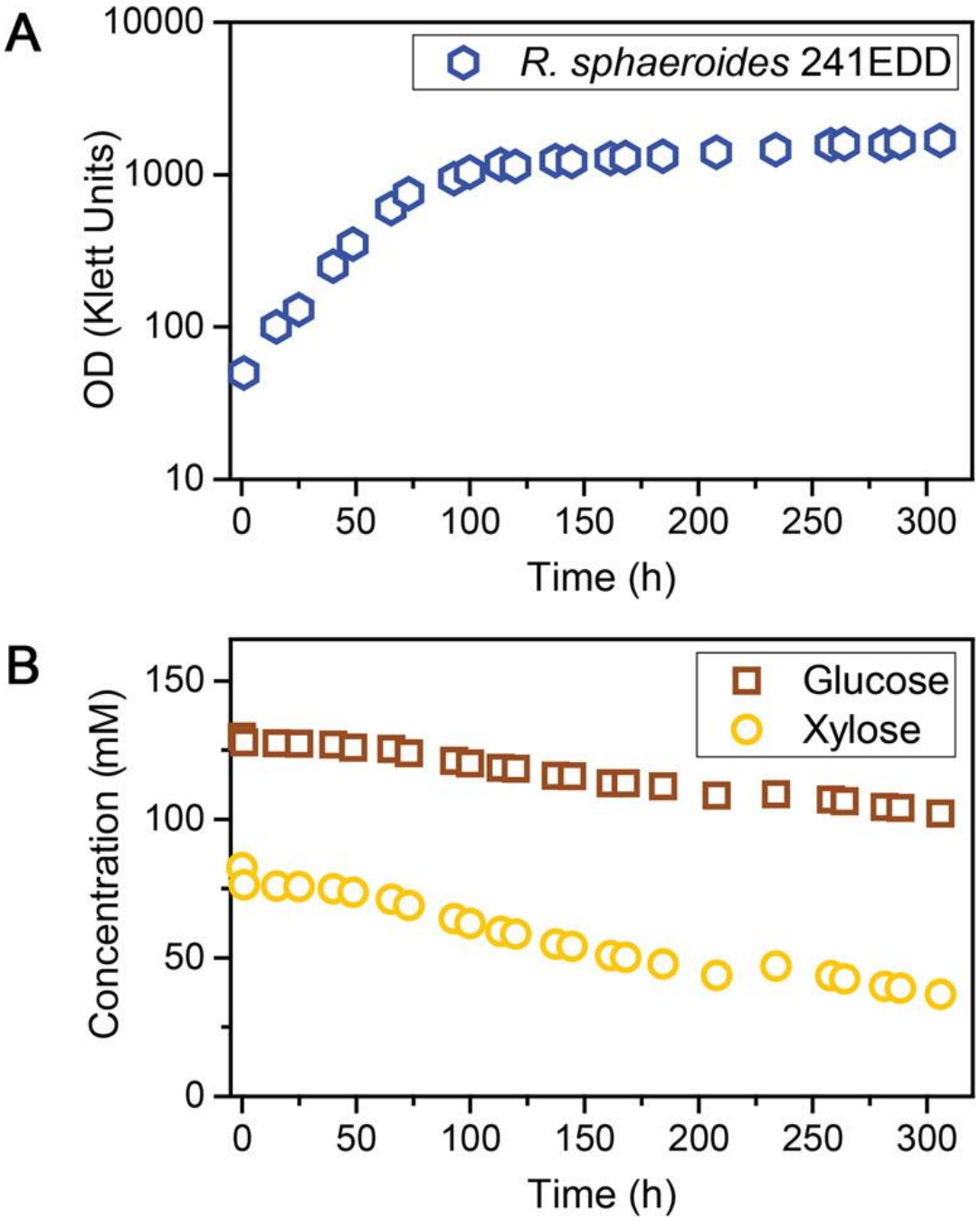


Figure 6.1.2.6. Growth of *R. sphaeroides* 241EDD (A) and sugar utilization (B) in ACSH that had been biologically conditioned with *R. palustris* CGA009. Note that *R. sphaeroides* 241EDD did not grow in ACSH that had not been conditioned with *R. palustris* CGA009.

Implications of the Biological Removal of Aromatics from Lignocellulosic Biomass Hydrolysates

Our results demonstrate the possibility to exploit *R. palustris* metabolism for removal of aromatic compounds from lignocellulosic biomass hydrolysates. A key observation is that *R. palustris* CGA009 grown in ACSH leaves the sugars unaltered and available for biofuel production by a second microbe because it preferentially uses acetate and aromatics as organic electron donors. Although demonstrated in this study using ACSH, removal of inhibitory aromatics with *R. palustris* could be applied to other biomass pretreatments where aromatics in the hydrolysates are also a concern.[138] In addition, the removal of acetate from ACSH by *R. palustris* could also provide added benefit to the use yeast and other microbes in which fermentation performance is inhibited by the presence of this organic acid in the biomass hydrolysate.[137,174] To achieve the full potential of this process, it is necessary to engineer strains capable of growing in more concentrated hydrolysates, and extend the range of plant-derived aromatics that can be metabolized. Moreover, by engineering strains capable of channeling the aromatics into specific phenolic compounds, as demonstrated here with several *R. palustris* mutants, it should be possible to both remove inhibitors and convert them to valuable bioproducts recoverable from the hydrolysates. The accumulation of well-defined phenolic compounds by engineered strains of *R. palustris* adds to the diversity of biochemicals that could be produced in a biorefinery,[175] and contributes to increasing the fraction of the carbon present in the hydrolysates that is recovered as a valuable product instead of being released as organic waste.

Supporting Information

Text describing the construction of *R. sphaeroides* 241EDD strain, additional figures supporting the reproducibility of aromatic degradation in ACSH by *R. palustris* CGA009 (Figure S1) and the

lack of growth of *R. sphaeroides* 241EDD in ACSH that has not been biologically conditioned by *R. palustris* CGA009 (Figure S2), and a table showing aromatic degradation in diluted ACSH **Table 6.1.2.2** The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b02062.

The authors declare no competing financial interest.

Acknowledgement

We thank Jackie Bastyr-Cooper for assistance performing analytical tests, and facilities at MSU and UW for the production of corn stover hydrolysates. This work was funded by the US Department of Energy Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). J. Zachary Oshlag was supported by a traineeship from the NIGMS Biotechnology Training grant (Grant T32 GM08349). We extend our thanks to Carrie S. Harwood for providing the *R. palustris* strains used in this study.

Supporting Information

A table taken from the supporting information is shown on the next page. More was included with the paper but only the most relevant material was included here. A complete collection of supplementary materials is available online as part of the open access publication at: http://pubs.acs.org/doi/abs/10.1021/acs.est.5b02062

Compound	Initial Concentration (µM)		Final Concentration (µM)*		Percent Removal**
	Average	Standard Deviation	Average	Standard Deviation	
Coumaroyl amide	606	16	0.0	0.0	100%
Feruloyl amide	302	3	0.5	0.1	100%
p-Coumaric acid	173	2	0.0	0.0	100%
Benzoic acid	51.8	1.2	0.0	0.0	100%
Vanillin	40.6	0.3	0.0	0.0	100%
Vanillamide	31.8	0.7	16.8	1.0	47%
4-Hydroxybenzoic acid	30.0	0.7	12.7	0.6	58%
Vanillic acid	27.6	0.7	0.0	0.0	100%
4-Hydroxybenzaldehyde	22.7	0.7	0.0	0.0	100%
Syringamide	20.7	1.1	23.8	0.2	-15%
Ferulic acid	18.3	0.8	0.0	0.0	100%
Syringic acid	8.5	0.2	11.5	0.7	-34%
4-Hydroxybenzamide	6.4	0.0	0.0	0.0	100%
Acetovanillone	3.8	0.0	7.8	0.3	-106%
Acetosyringone	3.5	0.0	3.2	0.2	7%
Syringaldehyde	2.7	0.0	0.0	0.0	100%
4-Hydroxyacetophenone	2.4	0.0	12.7	0.6	-440%
3,4-dihydroxybenzoic acid	1.3	0.0	225	22	-17852%
2,3-dihydroxybenzoic acid	0.3	0.0	0.1	0.2	58%
Sinapic acid	0.3	0.0	0.9	0.0	-195%

* Concentration after 14 days of incubation
** Negative values indicate increases in concentration

Table 6.1.2.2 Concentration of aromatic compounds in diluted ACSH before and after growth of *R. palustris* CGA009.

6.1.3 Formic-acid-induced depolymerization of oxidized lignin to aromatics

A.U. developed analytical methods, analyzed data, and produced graphics

This subsection has been published:

Rahimi, A., Ulbrich, A., Coon, J. J., & Stahl, S. S. (2014). Formic-acid-induced depolymerization of oxidized lignin to aromatics. *Nature*, 515(7526), 249-252

Abstract

Lignin is a heterogeneous aromatic biopolymer that accounts for nearly 30% of the organic carbon on Earth[176] and is one of the few renewable sources of aromatic chemicals.[115] As the most recalcitrant of the three components of lignocellulosic biomass (cellulose, hemicellulose and lignin)[116], lignin has been treated as a waste product in the pulp and paper industry, where it is burned to supply energy and recover pulping chemicals in the operation of paper mills.[177] Extraction of higher value from lignin is increasingly recognized as being crucial to the economic viability of integrated biorefineries.[124,125] Depolymerization is an important starting point for many lignin valorization strategies, because it could generate valuable aromatic chemicals and/or provide a source of low-molecular-mass feedstocks suitable for downstream processing.[127] Commercial precedents show that certain types of lignin (lignosulphonates) may be converted into vanillin and other marketable products[178,179], but new technologies are needed to enhance the lignin value chain. The complex, irregular structure of lignin complicates chemical conversion efforts, and known depolymerization methods typically afford ill-defined products in low yields (that is, less than 10-20wt%).[115,128,130] Here we describe a method for the depolymerization of oxidized lignin under mild conditions in aqueous formic acid that results in more than 60wt% yield of low-molecular-mass aromatics. We present the discovery of this facile C-O cleavage method, its application to aspen lignin depolymerization, and mechanistic insights into the reaction. The broader implications of these results for lignin conversion and biomass refining are also considered.

Main Text

Lignin is biosynthesized from a relatively small number of phenylpropanoid building blocks that contribute to recurring structural elements in this otherwise complex polymeric material. One of the most common motifs is the 'β-*O*-4' alkyl–aryl ether linkage between aromatic rings[115], which features a secondary benzylic alcohol at the Cα position and a primary aliphatic alcohol at the Cγ position (**Figure 6.1.3.1**). Fundamental studies of wood pulp delignification in the context of paper production have shown that oxidation of the Cα alcohol to a ketone (for example by using stoichiometric Mn or Cr oxide reagents) facilitates the removal of lignin from cellulose, apparently by promoting cleavage of the β-*O*-4 linkage.[180,181] This concept has not been applied to the production of low-molecular-mass aromatics from lignin. Several groups, including our own, have probed the reactions of oxidized lignin model compounds, such as **2** (**Figure 6.1.3.1b**), with varying degrees of success.[131,180,182-188] The recent development of a method for the chemoselective aerobic oxidation of Cα alcohols to ketones in native lignin, which achieved ~90% conversion to 'lignin^{ox}'[131], enables direct investigation of the reactivity of oxidized lignin. (Oxidative transformations of lignin and lignin model compounds have been the focus of extensive study and are summarized in refs 2, 20, 21.)

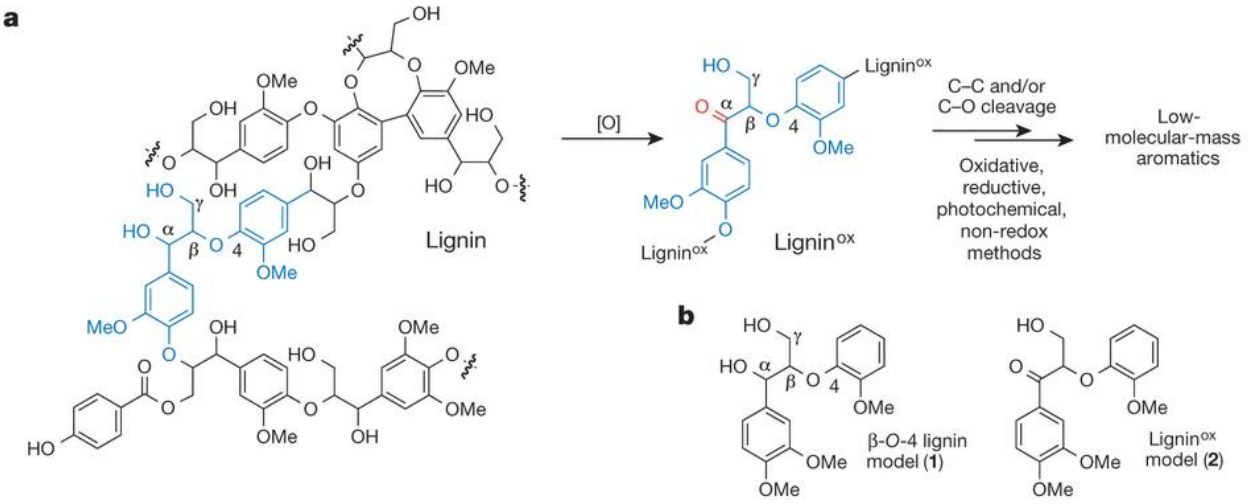
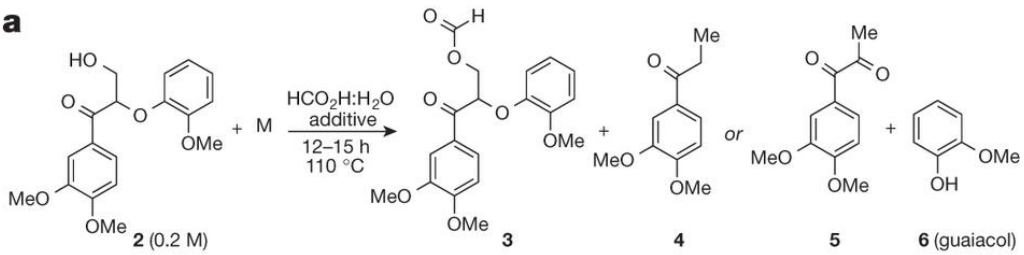


Figure 6.1.3.1 Lignin structure and strategies for depolymerization **a**, Representative structure of a fragment of poplar (including aspen) lignin, highlighting a β-*O*-4 unit, together with the strategy for lignin conversion to low-molecular-mass aromatics by chemoselective alcohol oxidation, followed by C–C and/or C–O cleavage. **b**, Structure of β-*O*-4 model compounds **1** and **2**.

We observed previously that the oxidized lignin model **2** reacts with alkaline hydrogen peroxide to yield the aromatic monomers veratric acid (88% yield) and guaiacol (42%).[131] The instability of the guaiacol under the reaction conditions prompted us to consider reductive cleavage methods. The reactivity of **2** was tested in the presence of different reducing metals, including zinc, aluminium, magnesium, iron and manganese, in aqueous formic acid at 110 °C (**Figure 6.1.3.2a**). Four different major products were identified from these reaction conditions. In the presence of zinc, **2** afforded small amounts of the *O*-formylated product **3** (6%), together with good yields of the aryl ethyl ketone reductive cleavage product **4** (76%) and guaiacol (69%) (**Figure 6.1.3.2a**, entry 1). When the reaction was performed with metal sources other than Zn (**Figure 6.1.3.2a**, entries 2–5), the diketone product **5** was observed instead of **4**. Good yields of **5** (74%) and guaiacol (63%) were obtained from the reaction with manganese. Numerous other conditions were tested with these metal sources; however, no further improvement in the product yields was observed. For example, substantially lower yields were obtained with organic solvents (ethanol, γ-valerolactone or toluene) or on replacement of formic acid with another acid source (acetic acid, HCl or H_2SO_4).

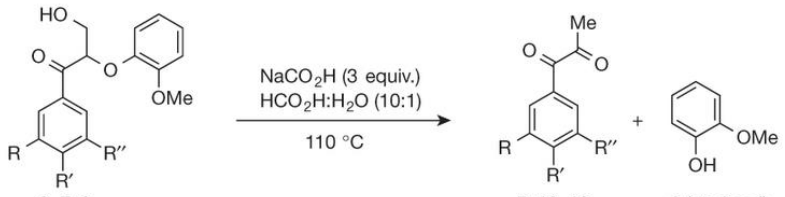
The conversion of **2** into the diketone product **5** is a redox-neutral process. We therefore tested the reaction in the absence of a reducing metal, and products **5** and **6** were obtained in yields comparable to the best results obtained with a stoichiometric metal source (**Figure 6.1.3.2a**, entry 6). Even better yields of **5** and **6** resulted (96% and 87%, respectively) when 3 equivalents of sodium formate were included in the reaction mixture (**Figure 6.1.3.2a**, entry 7). These conditions proved to be effective with several other lignin^{ox} model compounds, including those derived from *p*-hydroxyphenyl (H)-, guaiacyl (G)- and syringyl (S)-type lignin units **7–9** (**Figure 6.1.3.2b**). The Cα ketone is crucial to the success of the reaction: no cleavage products were observed when the non-oxidized model compound **1** was subjected to the formic acid/formate reaction conditions (**Figure 6.1.3.2c**).

a



Entry	M (2 equiv.)	Additive	3 (%)	4:5 (%)	6 (%)
1	Zn	–	6	76:0	69
2	Al	–	49	0:39	36
3	Mg	–	51	0:45	24
4	Fe	–	49	0:31	24
5	Mn	–	19	0:74	63
6	–	–	18	0:77	64
7	–	**HCO_2Na (3 equiv.)**	**0**	**0:96**	**87**

b



Lignin^{ox} model compound		Product: NMR yield (isolated yield)	
2: R = R′ = OMe, R″ = H	Non-phenolic model	**5**: 96% (92%)	**6**: 87% (81%)
7: R = R″ = H, R′ = OH	*p*-Hydroxyphenyl unit (H)	**10**: 93% (89%)	**6**: 82% (73%)
8: R = OMe, R′ = OH, R″ = H	Guaiacyl unit (G)	**11**: 92% (90%)	**6**: 89% (84%)
9: R = R″ = OMe, R′ = OH	Syringyl unit (S)	**12**: 93% (91%)	**6**: 90% (88%)

c

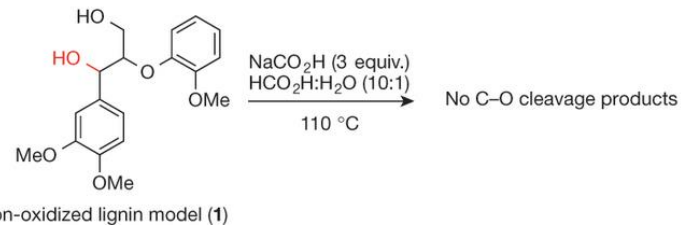


Figure 6.1.3.2 C-O cleavage of lignin model compounds with formic acid. **a**, Cleavage products obtained from the treatment of lignin^{ox} model compound **2** in formic acid with and without reducing metals (M). Yields shown in the table were determined by ^{1}H NMR spectroscopy, using hexamethyldisiloxane as the internal standard. **b**, Cleavage of different lignin^{ox} model compounds in the presence of formic acid and sodium formate. **c**, Control reaction with the non-oxidized lignin model compound **1** under the formic acid/formate cleavage conditions.

These results provided the basis for testing the reactivity of authentic lignin polymer. Aspen is a representative hardwood, and a sample of native aspen lignin was isolated and determined to have a ratio of S:G subunits of 2.2:1 on the basis of two-dimensional heteronuclear single quantum coherence NMR analysis.[189] After treatment of this material under the aerobic conditions described previously[131], the oxidized lignin was subjected to the formic acid/sodium formate reaction conditions at 110 °C. The amount of sodium formate added to the reaction mixture was estimated from the mass of the lignin sample to provide about 3 equivalents of formate per S/G aromatic subunit. The formic acid was evaporated after 24 h, and the residue was extracted with ethyl acetate. A soluble fraction, corresponding to 61.2wt% of the original lignin, was obtained, together with an insoluble fraction that accounted for 29.7wt% of the lignin (**Figure 6.1.3.3a**). Acetylation of the insoluble material and analysis by gel-permeation chromatography revealed a significant decrease in molecular mass relative to the initial lignin polymer (**Figure 6.1.3.3b**). The soluble fraction, containing the majority of the lignin-derived mass, was analysed by high-resolution mass spectrometry coupled with liquid chromatography (LC–MS) to identify and quantify the products of the reaction (**Figure 6.1.3.3 c** and **Supplementary Table 6.1.3.1**). Product identities were confirmed by comparison with authentic samples obtained commercially or synthesized independently.

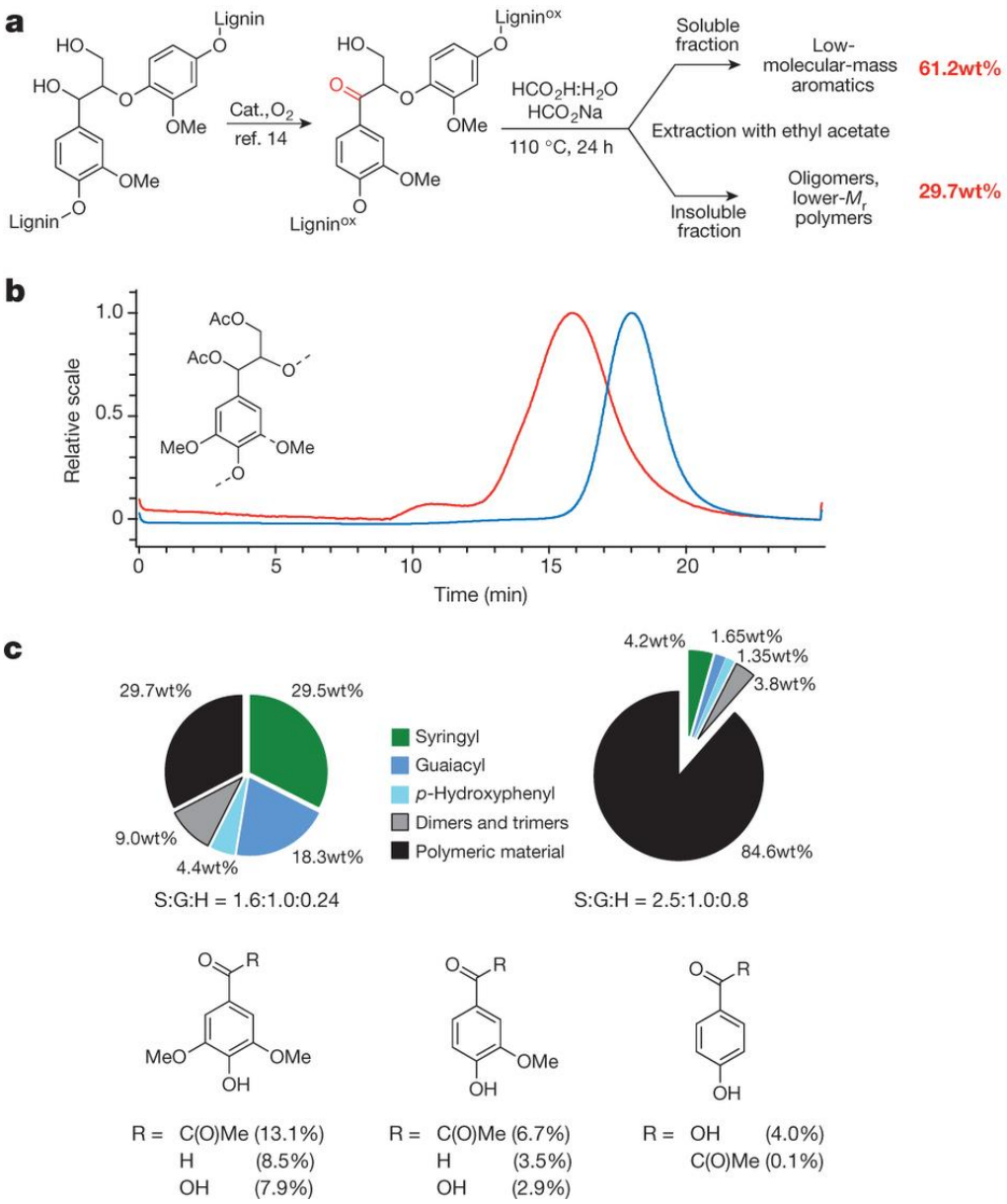


Figure 6.1.3.3 Depolymerization of aspen lignin with formic acid **a**, Depolymerization of oxidized aspen lignin. **b**, Gel-permeation chromatography of the insoluble fraction from depolymerization (29.7wt%) (blue trace) compared with that of native lignin (red trace). For the insoluble fraction of depolymerized lignin, M_w/M_n = 1.08 and M_n = 4,600 Da; for native lignin, M_w/M_n = 1.18 and M_n = 10,800 Da. **c**, Identification, quantification and distribution of products obtained from the depolymerization of oxidized lignin (upper left) compared with those from native lignin (upper right). Bottom: major depolymerization products from oxidized lignin. For a full product listing see **Supplementary Table 6.1.3.1**.

Overall, more than 52% of the original lignin was converted to well-defined aromatic compounds (**Figure 6.1.3.3c**). Syringyl and guaiacyl-derived diketones, directly analogous to those observed in the model study (compare **Figure 6.1.3.2b**), are two of the major depolymerization products (19.8%). The S:G diketone ratio (2:1) is very similar to the S:G monomer composition in the lignin (**Supplementary Figure 6.1.3.1**). Additional syringyl and guaiacyl-derived aromatics (for example, syringaldehyde and vanillin) account for most of the remaining product mass, together with *p*-hydroxybenzoic acid (4%), which is probably derived from *p*-hydroxybenzoic esters present in aspen (and other poplar) lignins.[190] A full listing of characterized products is provided in **Supplementary Table 6.1.3.1**.

An attempt to use unoxidized lignin in the reaction resulted in only 7.2wt% yield of low-molecular-mass aromatics. The products that are observed probably arise from the small amount of Cα ketones present in native lignin (evident from two-dimensional NMR spectroscopy). This result demonstrates the importance of Cα oxidation in promoting lignin depolymerization and provides strong motivation to develop improved lignin oxidation methods suitable for large-scale application. The Cα benzylic alcohol in lignin is activated electronically relative to the Cγ primary alcohol, and our previous study showed that numerous classes of oxidants promote the chemoselective oxidation of the Cα alcohol.[131] These observations bode well for the development of improved oxidation methods.

Mechanistic insights into the depolymerization process were obtained from additional studies of model compound **2**. Reaction time-course data, obtained by ^{1}H NMR spectroscopy, revealed that **2** converts quickly into the alcohol formylation product **3**, followed by a slower conversion of **3** into products **5** and **6** (**Figure 6.1.3.4a,b**). Conversion of **2** into **3** is rapid, even at room temperature (**Figure 6.1.3.4c**). Direct investigation of the reaction of **3** revealed a small steady-state concentration of the aryl vinyl ether intermediate **13** during the formation of diketone product **5** (**Figure 6.1.3.4d**). Independent testing of **13** confirmed that this intermediate is

consumed more rapidly than the overall conversion of **3** into **5** (**Figure 6.1.3.4e**). A large deuterium kinetic isotope effect, observed when the independent reaction rates of **2** and **2**-d_1 were compared (k_H/k_D = 9.3 ± 0.2; **Figure 6.1.3.4f**), indicates that the elimination of formic acid from **3** is the rate-limiting step in the transformation. This kinetic isotope effect is larger than the semi-classical limit associated with the cleavage of a C–H bond, but it aligns with a previous study providing evidence for proton tunnelling in concerted E2 elimination reactions.[191]

Overall, the sequence is a redox-neutral process that results in no net consumption of formic acid (**Figure 6.1.3.4g**). This feature distinguishes the present approach from other lignin conversion methods that employ formic acid as a source of H_2 in transfer hydrogenation/hydrogenolysis reactions with heterogeneous catalysts.[192,193] According to the mechanism in **Figure 6.1.3.4g**, the beneficial effect of lignin oxidation may be attributed to the ability of the benzylic carbonyl group to polarize the C–H bond and lower the barrier for the rate-limiting E2 elimination reaction. Both the increased C–H acidity and orbital overlap between the carbonyl C = O π system and the developing π bond of the alkene are expected to contribute to a lower barrier for this step. The mechanism also accounts for the beneficial effect of using a 'buffered' reaction medium, containing both formate and formic acid (compare **Figure 6.1.3.4a**, entries 6 and 7). The rate-limiting elimination step in **Figure 6.1.3.4g** is proposed to involve both a base (formate) to remove the proton and an acid (formic acid) to assist in the loss of the formate as a leaving group.

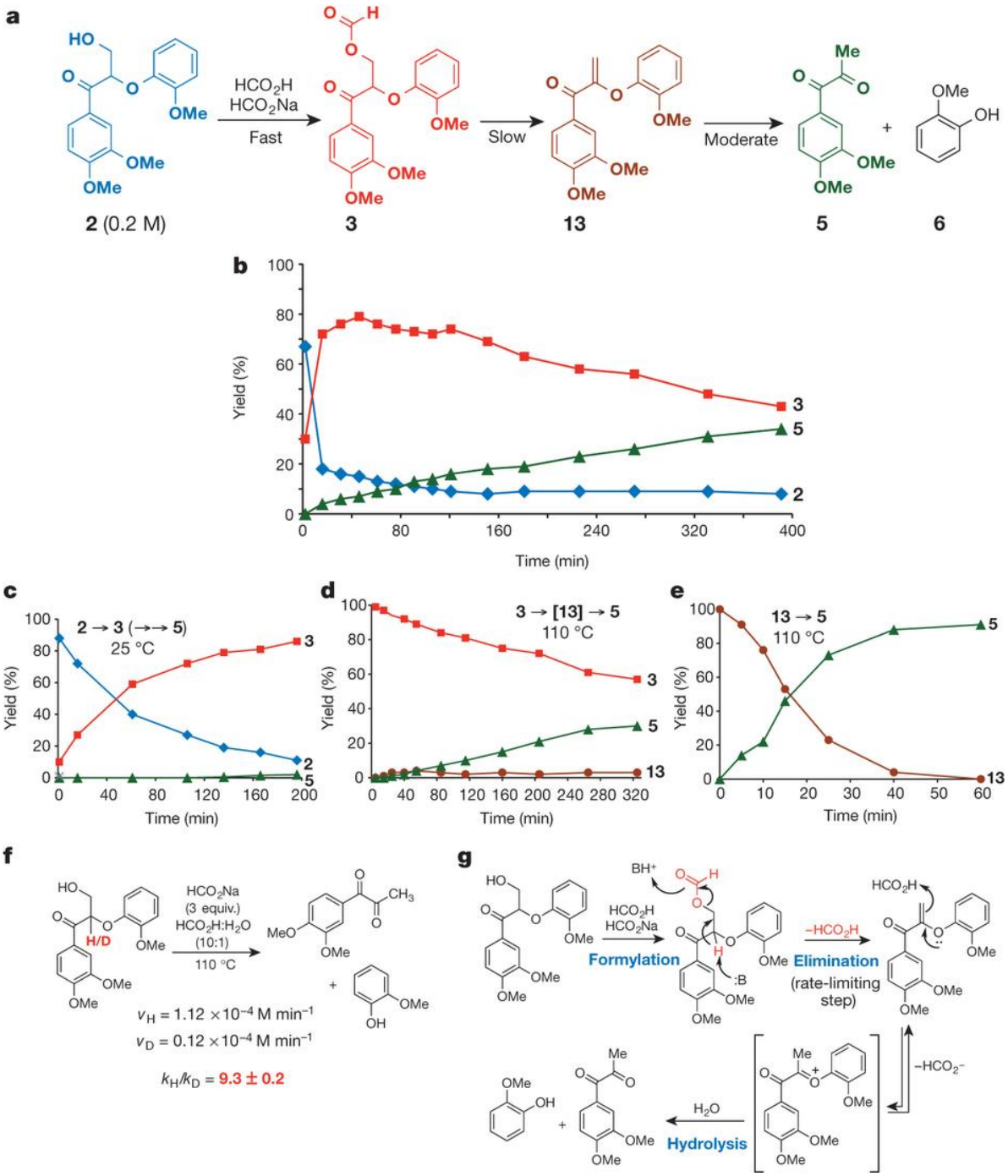


Figure 6.1.3.4 Mechanistic study of C-O cleavage in 2 **a**, **b**, Different intermediates (**a**) and overall time course of the reaction **2**→**3**→**[13]**→**5** at 110 °C (**b**). **c**–**e**, Time courses of formylation of **2** at 25 °C (**c**), elimination of formic acid from **3** at 110 °C (**d**), and hydrolysis of aryl vinyl ether **13** at 110 °C (**e**). **f**, Kinetic isotope effect for cleavage of the oxidized model compound **2**. **g**, Proposed mechanism for C–O cleavage.

The overall yield of structurally identified, monomeric aromatics obtained here is the highest reported so far for lignin depolymerization.[115,118,128,130] Although some of the products obtained from this process have direct commercial value (for example vanillin and syringaldehyde)[194,195], the more important result may be the generation of a stream of soluble aromatic feedstocks for further upgrading. For example, low-molecular-mass feedstocks should reduce coking and char formation, enhance conversions and facilitate product separations in processes with heterogeneous catalysts (such as hydrogenation, hydrogenolysis, decarbonylation and decarboxylation). The syringyl, guaiacyl and *p*-hydroxyphenyl aromatic products arise directly from the monomeric composition of the lignin, and their yields correlate closely with the quantity of β-*O*-4 linkages present in the original lignin. These observations highlight the importance of developing chemical conversion technologies for S-, G- and H-derived aromatics and suggest that plants containing lignin with high β-*O*-4 content (as much as 85% has been observed[196]) could be particularly appealing feedstocks for biomass valorization. This work further draws attention to biomass separation methods. Since the inception of the pulp and paper industry, biomass separation methods have emphasized the production of high-purity cellulose. Most existing process conditions, ranging from the classic kraft pulping to modern organosolvolysis methods, significantly modify or damage the lignin and lead to a significant decrease or complete loss of β-*O*-4 structural units. Biomass separation methods that deliver both high-purity sugar and native-type lignin streams, such as a recent γ-valerolactone-based process[197], could gain a competitive advantage in biorefining applications.

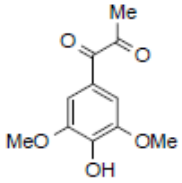
Supplementary Material

Supplementary figures and tables are enclosed below to augment the text. Not all supplementary material is shown due to its great length. A full collection of supplementary materials is available online at:

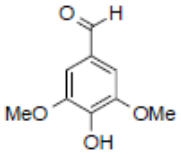
www.nature.com/nature/journal/v515/n7526/full/nature13867.html

Monomeric Products of Lignin Depolymerization		Yields (wt%) from Oxidized Lignin	Yields (wt%) from Native Lignin
Syringyl units	diketone I	13.1	2.4
	syringaldehyde	8.5	0.9
	syringic acid	7.9	0.9
Guaiacyl units	diketone II	6.7	1.0
	vanillin	3.5	0.3
	vanillic acid	2.9	0.3
	3,5-dinitroguaiacol	5.2	0
p-Hydroxyphenyl units	Diketone III	0.1	0
	4-hydroxy benzoic acid	4.0	1.3
	4-hydroxy-3-nitro benzoic acid	0.3	0
Total yield of characterized aromatic monomers		**52.2**	**7.1**
EtOAc soluble aromatic dimers and trimers		9.0	3.8
Total yield of soluble aromatics		**61.2**	**10.9**

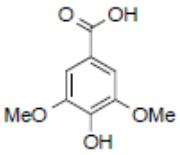
Chemical structures from syringyl units:



diketone I

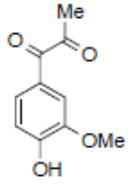


syringaldehyde

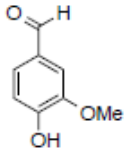


syringic acid

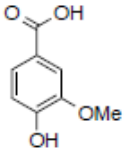
Chemical structures from guaiacyl units:



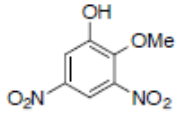
diketone II



vanillin

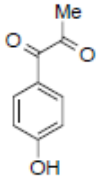


vanillic acid

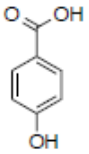


3,5-dinitroguaiacol

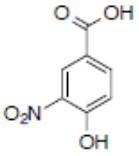
Chemical structures from p-hydroxyphenyl units:



diketone III

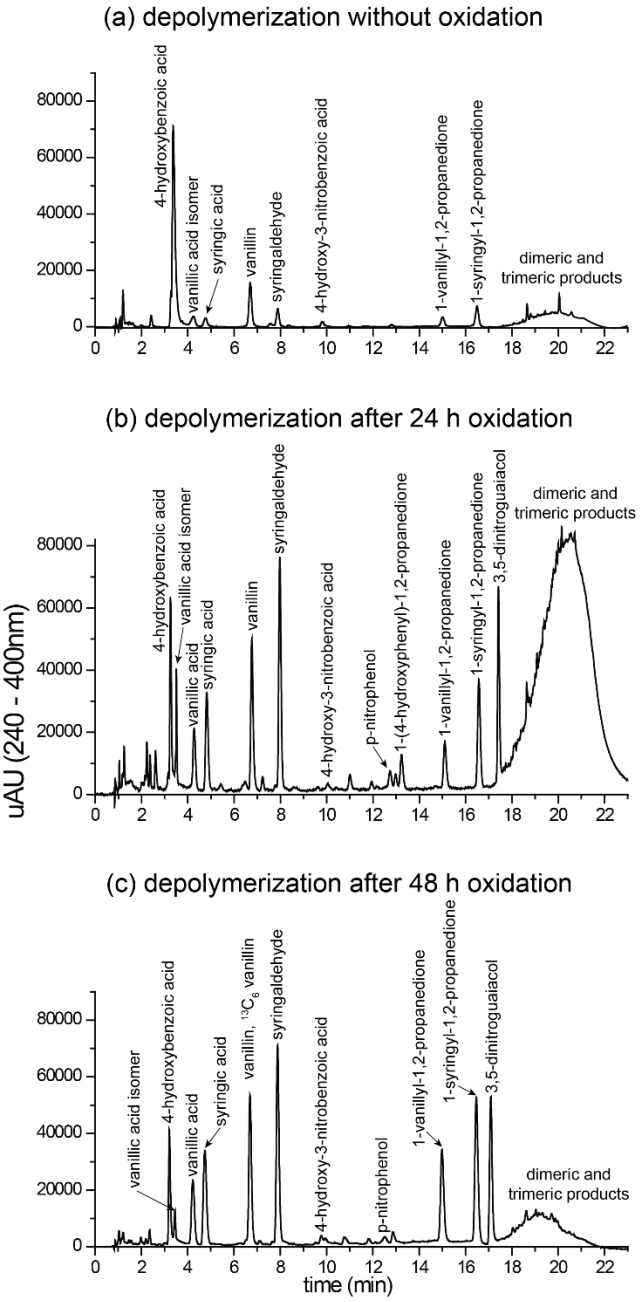


4-hydroxy benzoic acid



4-hydroxy-3-nitro benzoic acid

Supplementary Table 6.1.3.1 Identification of low molecular weight chemicals from oxidized and native lignin samples.



Supplementary Figure 6.1.3.1 Comparison of HPLC chromatograms from UV-traces. Main products are labeled. **a** depolymerization without oxidation. **b** depolymerization after 24 h oxidation **c** depolymerization after 48 h oxidation.

6.2: Methods and Applications for Coenzyme Q Biosynthesis Quantitation

Figures and results from the work presented in this sub-chapter have been published:

Lohman, D. C., Forouhar, F., Beebe, E. T., Stefely, M. S., Minogue, C. E., Ulbrich, A., ... & Pagliarini, D. J. (2014). Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis. *Proceedings of the National Academy of Sciences*, *111*(44), E4697-E4705.

Stefely, J. A., Reidenbach, A. G., Ulbrich, A., Oruganty, K., Floyd, B. J., Jochem, A., ... & Pagliarini, D. J. (2015). Mitochondrial ADCK3 employs an atypical protein kinase-like fold to enable coenzyme Q biosynthesis. *Molecular cell*, *57*(1), 83-94.

Abstract

Methods for quantitation of coenzyme Q (Q), its precursors, and other lipids are discussed in this chapter. Results obtained using targeted quantitation methods have contributed important pieces to the characterization of the formerly orphan mitochondrial proteins ADCK3 and COQ9. The work contributed to the identification of residues in these proteins that are critical to proper function of the coenzyme Q biosynthesis complex. Results from these studies were published recently.[18,19] For a large project interrogating the proteomes, metabolomes, and lipidomes of yeast strains harboring single protein knockouts a method quantifying 55 lipids across several classes was developed. Together with the other two omes, lipidomics contributed to novel assignments of functions to proteins. Mitochondrial HFD1 stands out since it was linked to the conversion of tyrosine into 4-hydrodrybenzoic acid, a precursor of coenzyme Q, for the first time through our work.

Introduction

The Coon Research Group has had a strong collaboration with Professor David Pagliarini's group in the biochemistry department at UW-Madison for several years.[22,38,198,199] Dr. Pagliarini's group focuses on mitochondrial biology with an emphasis on the characterization of orphan mitochondrial proteins.[200] Mitochondria are the power plants of eukaryotic cells. The majority of the ATP needed for homeostasis, motion, and growth is produced through mitochondrial respiration, i.e. the oxidation of nutrients to the more stable CO_2 molecule. The energy released from this process is then used to synthesize the high energy molecule ATP. Together with the citric acid cycle, the key pathway in a mitochondrion is the electron transport chain, the site of ATP synthesis driven by a proton gradient (**Figure 6.2.1**). A key non-protein member of the pathway is coenzyme Q, which carries electrons from complex I and complex II to complex III. The molecule was discovered at UW-Madison in 1957 and is a focal point of the Pagliarini Research Group.[201] Coenzyme Q deficiency obviously has serious consequences and the deficiency is not readily treated by Q supplementation.[202,203] Potential treatments must address

the source of the deficiency, defective coenzyme Q biosynthesis, which directly involves at least ten proteins with many more involved in pathways that deliver upstream precursors, such as 4-hydroxybenzaldehyde and isoprene units.[204] Final coenzyme Q biosynthesis proceeds from 4-hydroxybenzoic acid and a chain of prenyl units whose length varies from six to ten depending on the organism. After the joining of the polyprenyl tail to 4-hydroxybenzoic acid a series of enzymes catalyzes the stepwise decoration of the phenyl group to its final form. Steps exist in the pathway to which no protein has been assigned yet and proteins known to be obligatory for coenzyme Q production are yet to be assigned a specific function. Filling of these gaps is essential to provide therapeutic handles to treat certain coenzyme Q deficiencies. Patients may present with coenzyme Q deficiency yet harbor no known mutations related to the pathway. Treatment is thus much more difficult, if not impossible.

Two examples of proteins known to be important to Q biosynthesis yet with no known function are ADCK3 (COQ8) and COQ9. Through concerted efforts by biochemists, crystallographers, and LC-MS researchers that include the author's work, they have been assigned a more defined role in the pathway for the first time.[18,19] The measurements of levels of Q and some of its precursors in response to single amino acid substitutions in ADCK3 and COQ9 are described in this chapter.

In contrast to those two targeted in-depth protein characterizations, a large discovery study of nearly 200 single protein knockout yeast strains is underway as a collaboration between the Pagliarini and Coon Research Groups. Function of a previously uncharacterized protein is inferred from the similarities of the proteomes, metabolomes, and lipidomes of yeast lacking the uncharacterized protein to the omes of yeasts harboring known protein knockouts. Lipids across several classes including Q and precursors were monitored to provide a lipid profile for each strain. The lipidomics LC-MS method and one notable finding from the study are described in this chapter.

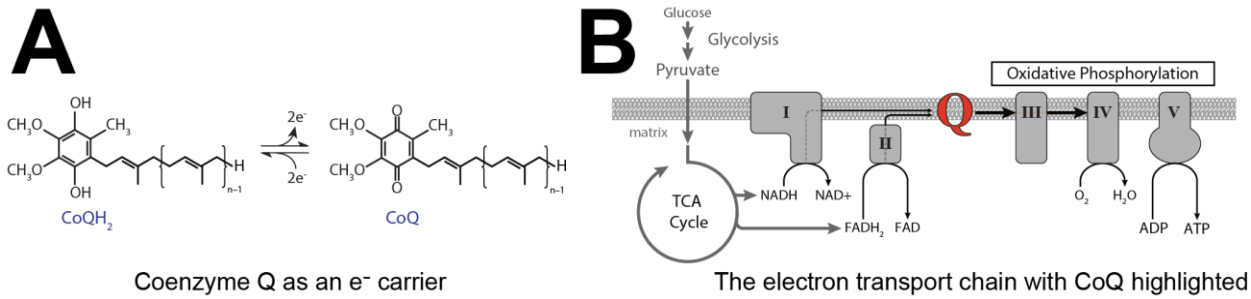


Figure 6.2.1 Coenzyme Q in the electron transport chain. **A** Coenzyme Q in its reduced quinol form and oxidized quinone forms. This redox activity allows the compound to shuttle electrons between complexes I&II and III. **B** The electron transport chain with coenzyme Q highlighted. This figure was adapted from a figure produced in the Pagliarini Research Group.

Methods

For the in-depth characterizations of COQ9 and ADCK3 two fast, targeted LC-MS and LC-MS/MS methods were developed. Detailed descriptions of both methods follow below. These descriptions were written by the author and have been published previously as stated on the first page of this chapter.[18,19] A description of the coenzyme Q quantitation method used for the ADCK3 study follows below:

> Yeast coenzyme Q (CoQ_6) quantitation by LC-MS $2.5x10^6$ yeast cells (as determined by OD_{600} of a starter culture) were used to inoculate a 25 mL culture of Ura– media (10 g/L glucose), which was incubated (30 °C, 230 rpm) for 23 h. At 23 h, the yeast cultures were ~4 h past the diauxic shift and the media was depleted of glucose. The OD_{600} of the culture was measured and used to determine the volume of culture needed to isolate $1x10^8$ yeast cells. $1x10^8$ yeast cells were pelleted by centrifugation (3,000 g, 3 min, 4 °C), the supernatant was discarded, and the yeast pellet was frozen at –20 °C. A frozen pellet of yeast (10^8 yeast cells) was thawed on ice and mixed with phosphate buffered saline (200 µL) and glass beads (0.5 mm diameter, 100 µL). The yeast were lysed by vortexing with the glass beads (30 s). Coenzyme Q10 (CoQ_{10}) was added as an internal standard (10 µM, 10 µL), and the lysate was vortexed (30 s). Hexanes/2-propanol (10:1, v/v) (500 µL) was added and vortexed (2 x 30 s). The samples were centrifuged (3,000 g, 1 min, 4 °C) to complete phase separation. 400 µL of the organic phase was transferred to a clean tube and dried under N_2(g). The organic residue was reconstituted in ACN/IPA/H_2O (65:30:5, v/v/v) (100 µL) by vortexing (30 s) and transferred to a glass vial for LC-MS analysis. LC-MS analysis was performed on an Ascentis Express C18 column (150 mm x 2.1 mm x 2.7 µm particle size, Supelco, Bellefonte, PA) using an Accela LC Pump (500 µL/min flow rate, Thermo Scientific, San Jose, CA). Mobile phase A consisted of 10 mM ammonium acetate in ACN/H_2O (70:30, v/v) containing 250 µL/L

acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. Initially, mobile phase B was held at 50% for 2 min and then increased to 95% over 3 min where it was held for 5 min. The column was re-equilibrated for 4 min before the next injection. Ten µL of sample were injected by an HTC PAL autosampler (Thermo Scientific, San Jose, CA). The LC system was coupled to a Q Exactive mass spectrometer (Build 2.3 SP2) by a HESI II heated ESI source kept at 350 °C (Thermo Scientific, San Jose, CA). The inlet capillary was kept at 350 °C, sheath gas was set to 60 units, and auxiliary gas to 15 units. The MS was operated in negative mode (2.5 kV) from 3 to 4.85 min with a mass range of 500–600 Th and an AGC target of $2x10^5$ and in positive mode (3 kV) from 4.7 to 7.5 min with a mass range spanning 550 to 900 Th and an AGC target of $1x10^6$. Resolving power was always set at 17,500. Quantitation was performed by integrating the peak areas of the $[M+H]^+$ ion of CoQ_6 at 591.44 Th and the $[M-H]^-$ ion of HAB_6 at 544.42 Th using the Xcalibur software suite (2.2 SP1.48, Thermo Scientific, San Jose, CA) and normalizing peak areas to the CoQ_{10} internal standard. Student's t-test was used to determine statistical significance.

Quantitation of coenzyme Q and hexaprenylaminobenzoic acid proceeded as follows for the COQ9 study. Due to lower signal from Q and precursors in this study, quantitation had to be adjusted to be done by tandem MS to increase selectivity. Sensitivity was also increased by the use of longer ion accumulation times of up to 500 ms. The description of the method that follows has been published previously and was written by the author.[18]

Yeast CoQ Quantitation by LC-MS. A total of 2.5×10^6 yeast cells (as determined by OD 600 of a starter culture) was used to inoculate a 25-mL culture of Ura – media (10 g/L glucose), which was incubated (30 °C, 230 rpm) for 23 h. After 23 h, the yeast cultures were ~ 4 h past the diauxic shift and the medium was depleted of glucose. The OD 600

of the culture was measured and used to determine the volume of culture needed to isolate 1×10^9 yeast cells. A total of 1×10^9 yeast cells was pelleted by centrifugation (3,000 × g, 3 min, 4 °C), the supernatant was discarded, and the yeast pellet was frozen at – 20 °C. Frozen pellets of yeast (2×10^9 cells total) were thawed on ice and mixed with PBS (400 µL), glass beads (0.5 mm diameter, 200 µL), and the yeast were lysed by vortexing with the glass beads (30 s). Q_{10} was added as an internal standard (10 µM, 10 µL). The yeast lysate was vortexed again (30 s). Hexanes/ 2-propanol (10:1, vol/vol) (1 mL) was added and vortexed (2 × 30 s). Brine (600 µL) was added and vortexed (2 × 30 s). The samples were centrifuged (3,000 × g, 5 min, 4 °C) to complete phase separation. A volume of 700 µL of the organic phase was transferred to a clean tube and dried under N_2(g) The organic residue was reconstituted in ACN/IPA/H_2O (65:30:5, vol/vol/vol) (100 µL) by vortexing (30 s) and transferred to a glass vial for LC-MS analysis. LC-MS analysis was performed on an Ascentis Express C18 column held at 35 °C (150 mm × 2.1 mm × 2.7 µm particle size; Supelco) using an Accela LC Pump (500 µL/min flow rate; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate in ACN/ H_2O (70:30, vol/vol) containing 250 µL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, vol/vol) with the same additives. Initially, mobile phase B was held at 50% for 1.5 min and then increased to 95% over 6.5 min where it was held for 2 min. The column was then reequilibrated for 3.5 min before the next injection. Ten microliters of sample were injected by an HTC PAL autosampler (Thermo Scientific). The LC system was coupled to a Q Exactive mass spectrometer (Build 2.3 SP2) by a HESI II heated ESI source kept at 325 °C (Thermo Scientific). The inlet capillary was kept at 350 °C, sheath gas was set to 60 units, and auxiliary gas to 20 units, and the spray voltage was set to 3,000 V. The MS was operated in scheduled targeted MS 2 (tMS^2) mode to quantify DMQ_6 and CoQ_6. From 4.5 to 6.25 min, the $[M + H]^+$ ions of DMQ_6 and CoQ_6 at 561.43 and 591.44 Th were isolated and fragmented, and

from 7.5 to 9.0 min, the $[M + H]^+$ ion of the internal standard CoQ_{10} at 880.72 Th was isolated and fragmented. Resolving power was set to 17,500, AGC target to 2×10^5 maximum injection time to 500 ms, isolation window to 1 Th, and collision energy to 27 units. Quantitation proceeded by integrating the peak areas of the characteristic product ions at 167.07 Th (DMQ_6) and 197.08 Th (CoQ species) using the Xcalibur software suite (2.2 SP1.48; Thermo Scientific) and normalizing to the internal standard. The separation was monitored for unusual peaks by UV detection at 265 nm (5 Hz, 49-nm filter bandwidth, 200-ms filter rise time) using an Accela PDA Detector (Thermo Scientific) set up in line with the MS.

For the larger discovery project a broader method was developed that targeted several representative species of phosphatidylcholines, phosphatidylethanolamines, phosphatidylinositols, phosphatidylserines, diacylglycerols, phosphatidic acids, lyso species, cardiolipins, and Q species. The LC-MS method was set up as described below. The text below is also part of a manuscript in preparation.

Lipids from 10 µL of extract were separated by LC on an Ascentis Express C18 column (150 mm x 2.1 mm x 2.7 µm particle size, Supelco, Bellefonte, PA) using an Accela HPLC pump (Thermo Fisher Scientific, San Jose, CA) at a flow-rate of 0.5 ml/min using a linear gradient. Mobile phase A was 70/30 acetonitrile/water containing 10mM ammonium acetate and 0.025 % acetic acid and B was 90/10 isopropanol/acetonitrile containing the same additives. Initially the flow was maintained at 50 % B for 1.5 min, then ramped to 95 % B over 7.5 min, held there for 2 min before returning to starting conditions over 0.5 min, and finally re-equilibrating the column for 2.5 min. The auto sampler (HTC PAL, Thermo Fisher Scientific) vigorously mixed each sample before injection to ensure homogeneity. MS conditions were as follows: a Q Exactive mass spectrometer (Thermo Scientific, Build 2.5) equipped with a HESI II spray source kept at

350 °C and +/- 4 kV was used for detection. The inlet capillary was kept at 350 °C, sheath gas was set to 60 units, and auxiliary gas to 20 units. Several scan functions were used to achieve optimal data acquisition for different lipid classes. For phospholipids, MS^1 data was acquired from 1 – 9 min at a resolving power of 35,000 with the AGC target set to $1x10^6$, mass range to 500 – 900 Th, and maximum injection time to 250 ms. For fatty acids and lyso species, MS^1 data was acquired from 0 – 3 min at a resolving power of 17,500 with the AGC target set to $5x10^5$, mass range to 220 – 600 Th, and maximum injection time to 100 ms. For cardiolipins, MS^1 data was acquired from 6.5 – 9.5 min at a resolving power of 17,500 with the AGC target set to $5x10^5$, mass range to 1320 – 1500 Th, and maximum injection time to 250 ms. For cytidine diacylglycerols, MS^1 data was acquired from 1 – 4.5 min at a resolving power of 17,500 with the AGC target set to $5x10^5$, mass range to 920 – 1050 Th, and maximum injection time to 250 ms. Quantitation for all of these species was performed by integrating the MS^1 peak areas of either the $[M-H]^-$ or $[M+Ac]^-$ ions. Coenzyme Q_6 and demethoxycoenzyme Q_6 were monitored from 4.7 to 5.8 min by tandem mass spectrometry using the 591.44 → 197.08 Th and 561.43 → 167.07 Th transitions at a normalized collision energy of 27 units, a resolving power of 17,500, a maximum injection time of 250 ms, and an isolation width of 1.5 Th. Peaks were automatically integrated using TraceFinder software (Thermo Fisher Scientific) and all integrations were checked manually.

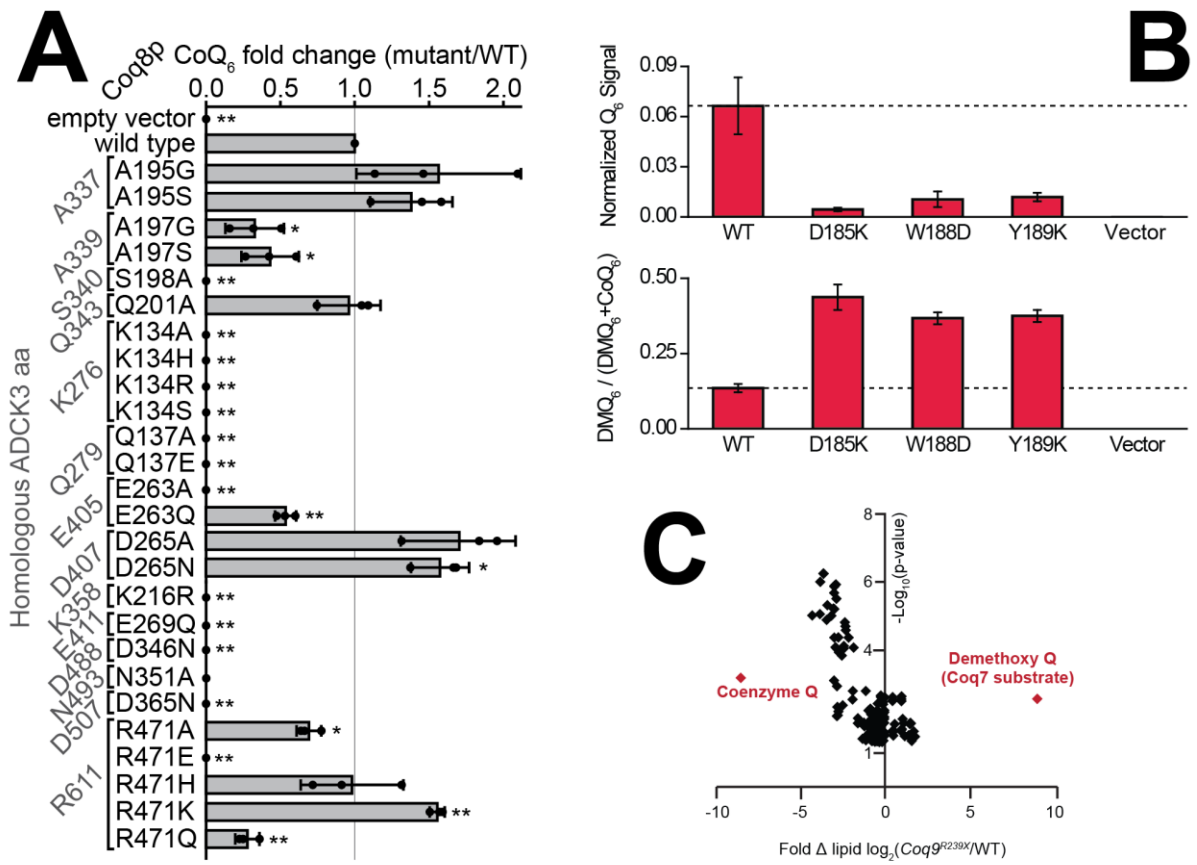


Figure 6.2.2 Effects of single amino acid mutations in ADCK3 and COQ9 proteins on Q levels. **A** Effects of several mutations of ADCK3 on Q levels in yeast. **B** Effects of three single amino acid substitutions in COQ9 on Q levels and DMQ buildup in yeast **C** Volcano plot showing the highly significant Q deficiency and buildup of DMQ, the COQ7 substrate, in mice with a mutated version of COQ9. Parts of this figure were adapted from figures produced in the Pagliarini group and have been published and presented before.[18,19,205]

Results

ADCK3 has been a puzzling protein since it shows sequence motifs typical of protein kinases, yet no in vitro kinase activity has ever been demonstrated.[206] We extended considerable efforts to try to show lipid kinase action but were not successful in doing so. Solving the protein's crystal structure, another part of the ADCK3 study, revealed that its tertiary structure is similar to protein kinase superfamily (PKL) members except that the substrate binding pocket is occluded. A number of single amino acid mutations predicted to affect structure and function were introduced in the Pagliarini lab and the lipid assay I developed measured their effect on coenzyme Q levels in the yeast organism. We found that several mutations either lead to a complete loss of function or severely depress Q levels (**Figure 6.2.2**). A single alanine to glycine (A197G) mutation stands out in that it also leads to a gain of kinase activity of the protein as evidenced by detection of phosphorylated peptides resulting from autophosphorylation. The proteomics experiment that identified the kinase activity was conducted by Catherine Minogue in the Coon laboratory. Identification of residues critical for functional Q synthesis, together with structural insights, allows for a better understanding of how ADCK3 mutations can cause severe neurodegenerative and other disease.[207]

COQ9 was shown to interact with COQ7 and to have a lipid binding site in the second study. These findings were the results of many different experiments ranging from the solving of the crystal structure to LC-MS proteomics.[18] Together with growth assays, key residues in COQ9 for protein function were identified by measuring Q levels and levels of precursors by LC-MS/MS. Residues 185, 188, and 189 are all located at the small molecule binding domain, indicating the importance of lipid binding for function. Results from the targeted MS/MS method were one key result in placing COQ9 in association with COQ7 at the step in the pathway that converts demethoxy Q (DMQ) into demethyl Q because a loss of function of COQ9 is accompanied by a buildup of $DMQ_{6,}$ the established substrate of COQ7(**Figure 6.3.2 B&C, Figure 6.2.3**).

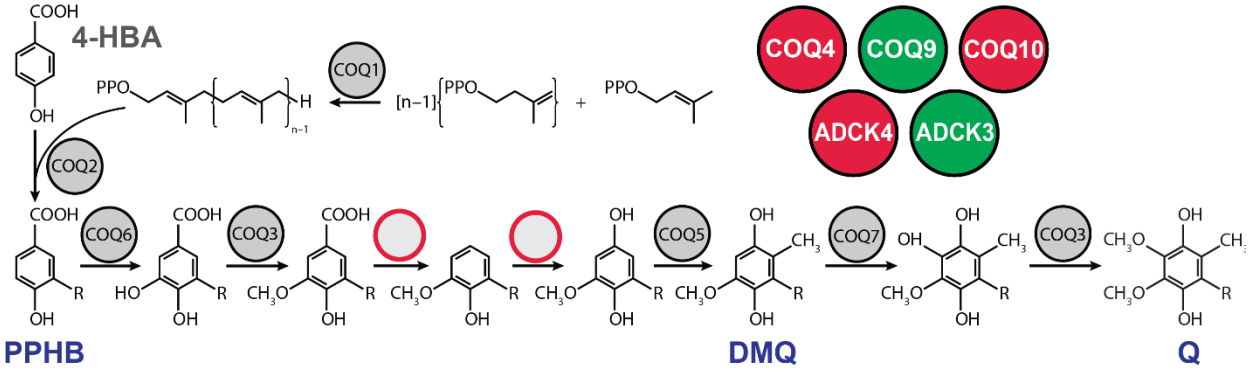


Figure 6.2.3 The coenzyme Q biosynthesis pathway. Species quantified by the assays described in this chapter are labeled with blue text. The enzymes catalyzing steps are given in grey circles. Several orphan mitochondrial proteins are shown on the top right. ADCK3 and COQ9 are labeled in green, since the knowledge of their structure and function was significantly expanded by the studies described in this chapter. This figure was adapted by the author from a similar figure produced in the Pagliarini Research Group.

The large discovery study has led to several significant new insights into mitochondrial proteins and a comprehensive manuscript is in preparation. Here, I highlight one example. We found that HFD1 knockout yeast can produce yeast only through an alternative pathway using 4-aminobenzoic acid instead of 4-hydroxybenzoic acid as the head group precursor.[208,209] This suggests that HFD1 is involved in an early step of coenzyme Q biosynthesis via 4-hydroxybenzoic acid. We identified this phenomenon by the lack of hexaprenylhydroxybenzoic ($PPHB_6$) acid, a Q precursor, in ΔHFD1 yeast (**Figure 6.2.4 A**). PPHB was one of the lipids targeted in the LC-MS assay developed for this project as was its amino analog hexaprenylaminobenzoic acid, which was present in the ΔHFD1 yeast. When the ΔHFD1 yeast were grown in media devoid of 4-hydroxy and 4-aminobezoic acid they did not produce any Q (**Figure 6.3.4 B)** and were incapable of growing on a non-fermentable carbon source (determined in separate growth assays done by Jonathan Stefely in the Pagliarini laboratory). Further studies of the behavior of ΔHFD1 are ongoing to more precisely determine HFD1's role in the mitochondria.

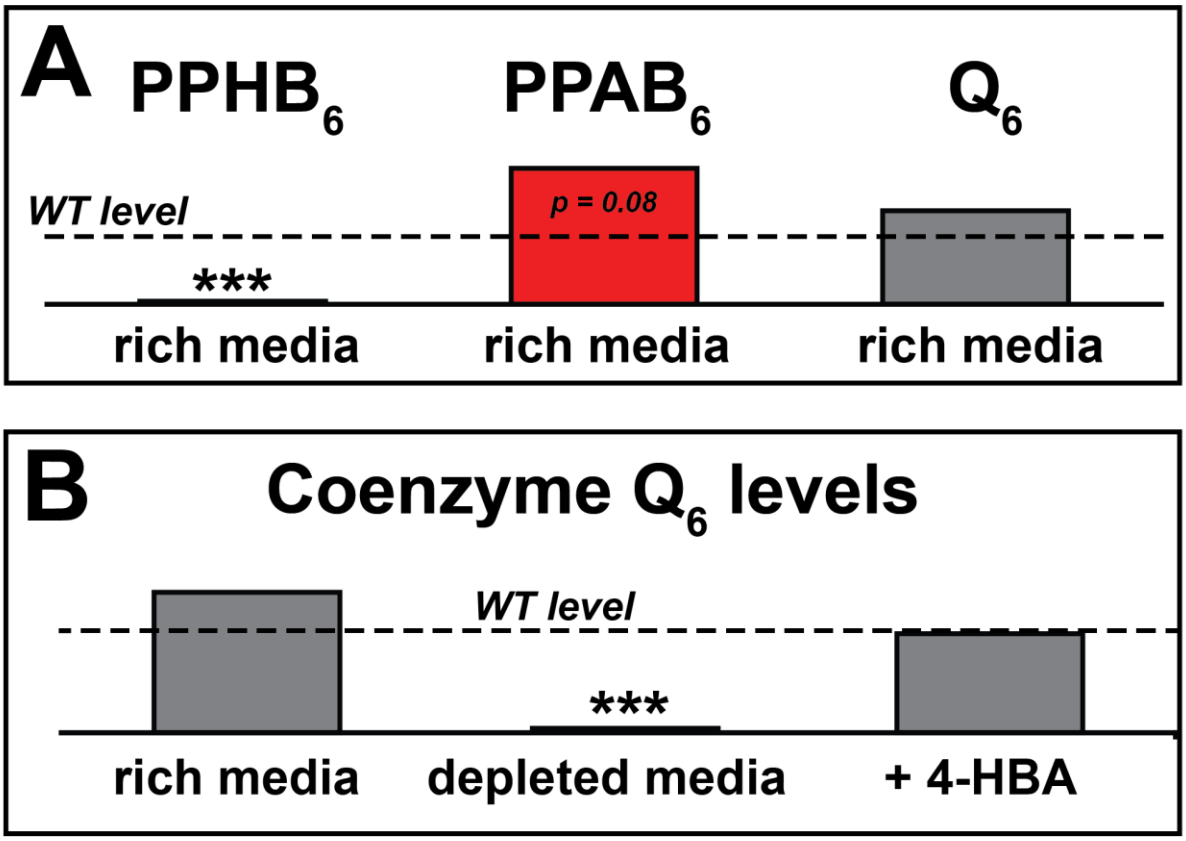


Figure 6.2.4 Changes in the levels of Q and precursors in ΔHFD1 yeast depending on growth media relative to wild type levels. **A** Media containing yeast extract that supplies both 4-hydroxybenzoic acid (4-HBA) and 4-aminobenzoic acid supports respiration in ΔHFD1 yeast. The precursor hexaprenylhydroxybenzoic ($PPHB_6$) acid is notably absent and only its amino analogue hexaprenylaminobenzoic acid is detected. The Coenzyme Q_6 level is not affected. **B** When ΔHFD1 yeast are grown in sparse synthetic media lacking 4-hydroxybenzoic and 4-aminobenzoic acid ΔHFD1 yeast cannot produce Q but respiration is rescued by 4-hydroxybenzoic acid supplementation. *** denotes $p < 0.001$.

References

(1) Patel, K. N.; Patel, J. K.; Patel, M. P.; Rajput, G. C.; Patel, H. A. *"Introduction to hyphenated techniques and their applications in pharmacy" Pharm Methods* 2010, *1*, 2-13.

(2) Harris, D. C. "*Quantitative chemical analysis"*, 8th ed.; W.H. Freeman and Co.: New York, 2010.

(3) Frenich, A. G.; Garcia, M. D. G.; Arrebola, F. J.; Vidal, J. L. M.; Galera, M. M.; Lopez, T. *"Determination of parts per trillion levels of benzoylurea pesticides in groundwater by high-performance liquid chromatography-electrospray ionization mass spectrometry" Chromatographia* 2000, *52*, 569-574.

(4) Skoog, D. A.; Holler, F. J.; Crouch, S. R. "*Principles of instrumental analysis"*, 6th ed.; Thomson Brooks/Cole: Belmont, CA, 2007, p xv, 1039 p.

(5) Pretsch, E.; Bühlmann, P.; Affolter, C. "*Structure determination of organic compounds : tables of spectral data"*, 3rd completely rev. and enl. English ed.; Springer: Berlin ; New York, 2000, p xv, 421 p.

(6) National Institute of Standards and Technology (U.S.). In *Secondary "NIST chemistry webbook"*; National Institute of Standards and Technology: Washington, D.C.

(7) Field, L. D.; Sternhell, S.; Kalman, J. R. "*Organic structures from spectra"*, 4th ed.; John Wiley and Sons Ltd.: Chichester ; Hoboken, N.J., 2008, p xiii, 453 p.

(8) Howard, G. A.; Martin, A. J. *"The separation of the C12-C18 fatty acids by reversed-phase partition chromatography" Biochem J* 1950, *46*, 532-8.

(9) Molnar, I.; Horvath, C. *"Reverse-Phase Chromatography of Polar Biological Substances - Separation of Catechol Compounds by High-Performance Liquid-Chromatography" Clin Chem* 1976, *22*, 1497-1502.

(10) James, A. T. *"Gas-liquid partition chromatography: the separation and identification of volatile bases; (a) aliphatic amines, (b) pyridine homologues" Biochem J* 1952, *51*, viii.

(11) McLafferty, F. W.; Ture*cek, F. "*Interpretation of mass spectra"*, 4th ed.; University Science Books: Mill Valley, Calif., 1993, p xviii, 371 p.

(12) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *"The Orbitrap: a new mass spectrometer" J Mass Spectrom* 2005, *40*, 430-43.

(13) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *"Electrospray ionization for mass spectrometry of large biomolecules" Science* 1989, *246*, 64-71.

(14) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. *"Protein and polymer analyses up to m/z 100 000 by laser ionization time? of? flight mass spectrometry" Rapid Commun Mass Sp* 1988, *2*, 151-153.

(15) Hoffmann, E. d.; Stroobant, V. "*Mass spectrometry : principles and applications"*, 3rd ed.; J. Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2007, p xii, 489 p.

(16) Ulbrich, A.; Merrill, A. E.; Hebert, A. S.; Westphall, M. S.; Keller, M. P.; Attie, A. D.; Coon, J. J. *"Neutron-Encoded Protein Quantification by Peptide Carbamylation"* *J Am Soc Mass Spectrom* 2013.

(17) Ulbrich, A.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. *"Organic acid quantitation by NeuCode methylamidation"* *Anal Chem* 2014, *86*, 4402-8.

(18) Lohman, D. C.; Forouhar, F.; Beebe, E. T.; Stefely, M. S.; Minogue, C. E.; Ulbrich, A.; Stefely, J. A.; Sukumar, S.; Luna-Sanchez, M.; Jochem, A.; Lew, S.; Seetharaman, J.; Xiao, R.; Wang, H.; Westphall, M. S.; Wrobel, R. L.; Everett, J. K.; Mitchell, J. C.; Lopez, L. C.; Coon, J. J.; Tong, L.; Pagliarini, D. J. *"Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis"* *Proc Natl Acad Sci U S A* 2014, *111*, E4697-705.

(19) Stefely, J. A.; Reidenbach, A. G.; Ulbrich, A.; Oruganty, K.; Floyd, B. J.; Jochem, A.; Saunders, J. M.; Johnson, I. E.; Minogue, C. E.; Wrobel, R. L.; Barber, G. E.; Lee, D.; Li, S.; Kannan, N.; Coon, J. J.; Bingman, C. A.; Pagliarini, D. J. *"Mitochondrial ADCK3 Employs an Atypical Protein Kinase-like Fold to Enable Coenzyme Q Biosynthesis"* *Mol Cell* 2015, *57*, 83-94.

(20) Rahimi, A.; Ulbrich, A.; Coon, J. J.; Stahl, S. S. *"Formic-acid-induced depolymerization of oxidized lignin to aromatics"* *Nature* 2014, *515*, 249-52.

(21) Sleno, L. *"The use of mass defect in modern mass spectrometry"* *J Mass Spectrom* 2012, *47*, 226-36.

(22) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. *"Neutron-encoded mass signatures for multiplexed proteome quantification"* *Nat Methods* 2013.

(23) Kind, T.; Fiehn, O. *"Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm"* *Bmc Bioinformatics* 2006, *7*.

(24) Nielsen, S. S. "*Food analysis*", 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2003, p 557 p.

(25) Sanz-Vicente, I.; Cabredo, S.; Galban, J. *"Gas chromatography with UV-Vis molecular absorption spectrometry detection: Increasing sensitivity of the determination of alcohols and phenols by derivatization"* *Chromatographia* 1998, *48*, 542-547.

(26) Skoog, D. A. "*Fundamentals of analytical chemistry*", 9th Ed. ed.; Cengage - Brooks/Cole: Belmont, CA, 2012.

(27) Woodward, R. B. *"Structure and the absorption spectra of α, β-unsaturated ketones"* *Journal of the American Chemical Society* 1941, *63*, 1123-1126.

(28) Austin, S.; Kontur, W. S.; Ulbrich, A.; Oshlag, J. Z.; Zhang, W.; Higbee, A.; Zhang, Y.; Coon, J. J.; Hodge, D. B.; Donohue, T. J.; Noguera, D. R. *"Metabolism of Multiple Aromatic Compounds in Corn Stover Hydrolysate by Rhodopseudomonas palustris"* *Environ Sci Technol* 2015.

(29) Keating, D. H.; Zhang, Y.; Ong, I. M.; McIlwain, S.; Morales, E. H.; Grass, J. A.; Tremaine, M.; Bothfeld, W.; Higbee, A.; Ulbrich, A.; Balloon, A. J.; Westphall, M. S.; Aldrich, J.; Lipton, M. S.; Kim, J.; Moskvin, O. V.; Bukhman, Y. V.; Coon, J. J.; Kiley, P. J.; Bates, D. M.; Landick, R. *"Aromatic inhibitors derived from*

ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification" *Front Microbiol* 2014, *5*, 402.

(30) Greenbaum, D.; Colangelo, C.; Williams, K.; Gerstein, M. *"Comparing protein abundance and mRNA expression levels on a genomic scale"* *Genome Biol* 2003, *4*.

(31) Zubarev, R. A. *"The challenge of the proteome dynamic range and its implications for in-depth proteomics"* *Proteomics* 2013, *13*, 723-726.

(32) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *"An automated multidimensional protein identification technology for shotgun proteomics"* *Anal Chem* 2001, *73*, 5683-5690.

(33) Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. *"The dynamic range of protein expression: A challenge for proteomic research"* *Electrophoresis* 2000, *21*, 1104-1115.

(34) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. *"Comparison of label-free methods for quantifying human proteins by shotgun proteomics"* *Molecular & Cellular Proteomics* 2005, *4*, 1487-1502.

(35) Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chen, K.; Wahl, R. C.; Yu, W.; Luethy, R.; Patterson, S. D. *"Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I. Profiling an unfractionated tryptic digest"* *Proteomics* 2001, *1*, 93-107.

(36) Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood, J.; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R. *"Tools for glycomics: relative quantitation of glycans by isotopic permethylation using (CH3I)-C-13"* *Glycobiology* 2007, *17*, 677-687.

(37) Vincent, C. E.; Potts, G. K.; Ulbrich, A.; Westphall, M. S.; Atwood, J. A., 3rd; Coon, J. J.; Weatherly, D. B. *"Segmentation of precursor mass range using "tiling" approach increases peptide identifications for MS1-based label-free quantification"* *Anal Chem* 2013, *85*, 2825-32.

(38) Hebert, A. S.; Dittenhafer-Reed, K. E.; Yu, W.; Bailey, D. J.; Selen, E. S.; Boersma, M. D.; Carson, J. J.; Tonelli, M.; Balloon, A. J.; Higbee, A. J.; Westphall, M. S.; Pagliarini, D. J.; Prolla, T. A.; Assadi-Porter, F.; Roy, S.; Denu, J. M.; Coon, J. J. *"Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome"* *Mol Cell* 2013, *49*, 186-199.

(39) Angel, P. M.; Orlando, R. *"Quantitative carbamylation as a stable isotopic labeling method for comparative proteomics"* *Rapid Commun Mass Sp* 2007, *21*, 1623-1634.

(40) Elias, J. E.; Gygi, S. P. *"Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry"* *Nat Methods* 2007, *4*, 207-14.

(41) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *"Open mass spectrometry search algorithm"* *J Proteome Res* 2004, *3*, 958-64.

(42) Wenger, C. D.; Phanstiel, D. H.; Lee, M. V.; Bailey, D. J.; Coon, J. J. *"COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA"* *Proteomics* 2011, *11*, 1064-74.

(43) Denisov, E.; Damoc, E.; Lange, O.; Makarov, A. *"Orbitrap mass spectrometry with resolving powers above 1,000,000" Int J Mass Spectrom* 2012, *325*, 80-85.

(44) Swaney, D. L.; Wenger, C. D.; Coon, J. J. *"Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics" J Proteome Res* 2010, *9*, 1323-1329.

(45) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *"Quantitative mass spectrometry in proteomics: a critical review" Anal Bioanal Chem* 2007, *389*, 1017-1031.

(46) Kozawa, S.; Honda, A.; Kajiwara, N.; Takemoto, Y.; Nagase, T.; Nikami, H.; Okano, Y.; Nakashima, S.; Shimozawa, N. *"Induction of peroxisomal lipid metabolism in mice fed a high-fat diet" Mol Med Rep* 2011, *4*, 1157-1162.

(47) Sparks, L. M.; Xie, H.; Koza, R. A.; Mynatt, R.; Hulver, M. W.; Bray, G. A.; Smith, S. R. *"A high-fat diet coordinately downregulates genes required for mitochondrial oxidative phosphorylation in skeletal muscle" Diabetes* 2005, *54*, 1926-1933.

(48) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *"Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources" Nat Protoc* 2009, *4*, 44-57.

(49) Kasper, D. C.; Ratschmann, R.; Metz, T. F.; Mechtler, T. P.; Moslinger, D.; Konstantopoulou, V.; Item, C. B.; Pollak, A.; Herkner, K. R. *"The National Austrian Newborn Screening Program - Eight years experience with mass spectrometry. Past, present, and future goals" Wien Klin Wochenschr* 2010, *122*, 607-613.

(50) Kostiainen, R.; Kotiaho, T.; Kuuranne, T.; Auriola, S. *"Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies" Journal of Mass Spectrometry* 2003, *38*, 357-372.

(51) Maurer, H. H. *"Liquid chromatography mass spectrometry in forensic and clinical toxicology" J Chromatogr B* 1998, *713*, 3-25.

(52) Mensch, J.; Noppe, M.; Adriaensen, J.; Melis, A.; Mackie, C.; Augustijns, P.; Brewster, M. E. *"Novel generic UPLC/MS/MS method for high throughput analysis applied to permeability assessment in early Drug Discovery" J Chromatogr B* 2007, *847*, 182-187.

(53) Peng, J. M.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *"Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome" J Proteome Res* 2003, *2*, 43-50.

(54) Aebersold, R.; Mann, M. *"Mass spectrometry-based proteomics" Nature* 2003, *422*, 198-207.

(55) Careri, M.; Bianchi, F.; Corradini, C. *"Recent advances in the application of mass spectrometry in food-related analysis" J Chromatogr A* 2002, *970*, 3-64.

(56) Herderich, M. *"Mass spectrometry techniques in food analysis" Nachr Chem* 2001, *49*, 378-381.

(57) Mastovska, K. *"Multiresidue Analysis of Antibiotics in Food of Animal Origin Using Liquid Chromatography-Mass Spectrometry" Methods Mol Biol* 2011, *747*, 267-307.

(58) Pardo, O.; Yusa, V.; Leon, N.; Pastor, A. *"Development of a method for the analysis of seven banned azo-dyes in chilli and hot chilli food samples by pressurised liquid extraction and liquid chromatography with electrospray ionization-tandem mass spectrometry" Talanta* 2009, *78*, 178-186.

(59) Pico, Y.; Barcelo, D. *"The expanding role of LC-MS in analyzing metabolites and degradation products of food contaminants" Trac-Trend Anal Chem* 2008, *27*, 821-835.

(60) Kopka, J. *"Current challenges and developments in GC-MS based metabolite profiling technology" J Biotechnol* 2006, *124*, 312-322.

(61) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *"Accurate quantitation of protein expression and site-specific phosphorylation" Proc Natl Acad Sci U S A* 1999, *96*, 6591-6.

(62) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. *"Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS" Anal Chem* 2003, *75*, 1895-904.

(63) Christoforou, A. L.; Lilley, K. S. *"Isobaric tagging approaches in quantitative proteomics: the ups and downs" Anal Bioanal Chem* 2012, *404*, 1029-37.

(64) Merrill, A. E.; Coon, J. J. *"Quantifying proteomes and their post-translational modifications by stable isotope label-based mass spectrometry" Current opinion in chemical biology* 2013.

(65) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *"Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics" Mol Cell Proteomics* 2002, *1*, 376-86.

(66) Guo, K.; Li, L. *"Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome" Anal Chem* 2009, *81*, 3919-3932.

(67) Lamos, S. M.; Shortreed, M. R.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. *"Relative quantification of carboxylic acid metabolites by liquid chromatography - Mass spectrometry using isotopic variants of cholamine" Anal Chem* 2007, *79*, 5143-5149.

(68) Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *"In vitro stable isotope labeling for discovery of novel metabolites by liquid chromatography-mass spectrometry: Confirmation of gamma-tocopherol metabolism in human A549 cell" J Chromatogr A* 2010, *1217*, 667-75.

(69) Hebert, A. S.; Merrill, A. E.; Stefely, J. A.; Bailey, D. J.; Wenger, C. D.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. *"Amine-reactive neutron-encoded labels for highly-plexed proteomic quantitation" Molecular & Cellular Proteomics* 2013.

(70) Richards, A. L.; Vincent, C. E.; Guthals, A.; Rose, C. M.; Westphall, M. S.; Bandeira, N.; Coon, J. J. *"Neutron-encoded signatures enable automated product ion annotation from tandem mass spectra" Mol Cell Proteomics* 2013.

(71) Rose, C. M.; Merrill, A. E.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. *"Neutron Encoded Labeling for Peptide Identification" Anal Chem* 2013, *85*, 5129-5137.

(72) American Oil Chemists' Society.; Firestone, D. "*Official methods and recommended practices of the AOCS"*, 6th ed.; AOCS: Urbana, Ill., 2009, p v. (loose-leaf).

(73) Aveldano, M. I.; Horrocks, L. A. *"Quantitative Release of Fatty-Acids from Lipids by a Simple Hydrolysis Procedure" Journal of lipid research* 1983, *24*, 1101-1105.

(74) Knoess, H. P.; Neeland, E. G. *"A modified synthesis of the insect repellent DEET" J Chem Educ* 1998, *75*, 1267-1268.

(75) Council, I. O., "Trade Standard Applying to Olive Oils and Olive-Pomace Oils"; International Olive Council Madrid, 2012.

(76) Peterson, A. C.; Hauschild, J. P.; Quarmby, S. T.; Krumwiede, D.; Lange, O.; Lemke, R. A. S.; Grosse-Coosmann, F.; Horning, S.; Donohue, T. J.; Westphall, M. S.; Coon, J. J.; Griep-Raming, J. *"Development of a GC/Quadrupole-Orbitrap Mass Spectrometer, Part I: Design and Characterization" Anal Chem* 2014, *86*, 10036-10043.

(77) Patel, V. J.; Thalassinos, K.; Slade, S. E.; Connolly, J. B.; Crombie, A.; Murrell, J. C.; Scrivens, J. H. *"A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches" J Proteome Res* 2009, *8*, 3752-3759.

(78) Zhu, W. H.; Smith, J. W.; Huang, C. M. *"Mass Spectrometry-Based Label-Free Quantitative Proteomics" J Biomed Biotechnol* 2010.

(79) Wang, G. H.; Wu, W. W.; Zeng, W. H.; Chou, C. L.; Shen, R. F. *"Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes" J Proteome Res* 2006, *5*, 1214-1223.

(80) Chelius, D.; Bondarenko, P. V. *"Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry" J Proteome Res* 2002, *1*, 317-323.

(81) Wang, W. X.; Zhou, H. H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *"Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards" Anal Chem* 2003, *75*, 4818-4826.

(82) Higgs Jr, R. E.; Julian Jr, R. K.; Kaiser Jr, R. E. In *Secondary "System and methods for qualitatively and quantitatively comparing complex admixtures using single ion chromatograms derived from spectroscopic analysis of such admixtures"*; Google Patents, 1999.

(83) Chelius, D.; Zhang, T.; Wang, G. H.; Shen, R. F. *"Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry" Anal Chem* 2003, *75*, 6658-6665.

(84) Bondarenko, P. V.; Chelius, D.; Shaler, T. A. *"Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry" Anal Chem* 2002, *74*, 4741-4749.

(85) Finney, G. L.; Blackler, A. R.; Hoopmann, M. R.; Canterbury, J. D.; Wu, C. C.; MacCoss, M. J. *"Label-free comparative analysis of proteomics mixtures using chromatographic alignment of high-resolution mu LC-MS data" Anal Chem* 2008, *80*, 961-971.

(86) Higgs, R. E.; Knierman, M. D.; Gelfanova, V.; Butler, J. P.; Hale, J. E. *"Comprehensive label-free method for the relative quantification of proteins from biological samples" J Proteome Res* 2005, *4*, 1442-1450.

(87) Norton, S. M. In *Secondary "Methods for time-alignment of liquid chromatography-mass spectrometry data"*; Google Patents, 2009.

(88) Panchaud, A.; Affolter, M.; Moreillon, P.; Kussmann, M. *"Experimental and computational approaches to quantitative proteomics: Status quo and outlook" J Proteomics* 2008, *71*, 19-33.

(89) Radulovic, D.; Jelveh, S.; Ryu, S.; Hamilton, T. G.; Foss, E.; Mao, Y. Y.; Emili, A. *"Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry" Molecular & Cellular Proteomics* 2004, *3*, 984-997.

(90) Dowell, J. A.; Frost, D. C.; Zhang, J.; Li, L. J. *"Comparison of two-dimensional fractionation techniques for shotgun proteomics" Anal Chem* 2008, *80*, 6715-6723.

(91) Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. *"Multidimensional separations-based shotgun proteomics" Chem Rev* 2007, *107*, 3654-3686.

(92) Hanash, S. M. *"Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: Current status" Electrophoresis* 2000, *21*, 1202-1209.

(93) Villen, J.; Gygi, S. P. *"The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry" Nat Protoc* 2008, *3*, 1630-1638.

(94) Wang, H.; Chang-Wong, T.; Tang, H. Y.; Speicher, D. W. *"Comparison of Extensive Protein Fractionation and Repetitive LC-MS/MS Analyses on Depth of Analysis for Complex Proteomes" J Proteome Res* 2010, *9*, 1032-1040.

(95) Luber, C. A.; Cox, J.; Lauterbach, H.; Fancke, B.; Selbach, M.; Tschopp, J.; Akira, S.; Wiegand, M.; Hochrein, H.; O'Keeffe, M.; Mann, M. *"Quantitative Proteomics Reveals Subset-Specific Viral Recognition in Dendritic Cells" Immunity* 2010, *32*, 279-289.

(96) Altelaar, A. F. M.; Heck, A. J. R. *"Trends in ultrasensitive proteomics" Current Opinion in Chemical Biology* 2012, *16*, 206-213.

(97) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Frohlich, F.; Cox, J.; Mann, M. *"Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation" Molecular & Cellular Proteomics* 2011, *10*.

(98) Washburn, M. P.; Wolters, D.; Yates, J. R. *"Large-scale analysis of the yeast proteome by multidimensional protein identification technology" Nat Biotechnol* 2001, *19*, 242-247.

(99) Davis, M. T.; Spahr, C. S.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Yu, W.; Luethy, R.; Patterson, S. D. *"Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry - II. Limitations of complex mixture analyses" Proteomics* 2001, *1*, 108-117.

(100) Yi, E. C.; Marelli, M.; Lee, H.; Purvine, S. O.; Aebersold, R.; Aitchison, J. D.; Goodlett, D. R. *"Approaching complete peroxisome characterization by gas-phase fractionation" Electrophoresis* 2002, *23*, 3205-3216.

(101) Scherl, A.; Shaffer, S. A.; Taylor, G. K.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. *"Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides" Anal Chem* 2008, *80*, 1182-1191.

(102) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *"The one hour yeast proteome" Mol Cell Proteomics* 2014, *13*, 339-47.

(103) Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *"One-hour proteome analysis in yeast" Nat. Protocols* 2015, *10*, 701-714.

(104) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q. Y.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; Bui, H.; Specht, A.; Lange, O.; Denisov, E.; Makarov, A.; Horning, S.; Zabrouskov, V. *"Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates" Anal Chem* 2013, *85*, 11710-11714.

(105) Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A. *"A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics" Anal Bioanal Chem* 2003, *376*, 952-965.

(106) Le Blanc, J. C. Y.; Hager, J. W.; Ilisiu, A. M. P.; Hunter, C.; Zhong, F.; Chu, I. *"Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications" Proteomics* 2003, *3*, 859-869.

(107) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Francesco, V. D.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R.-R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z. Y.; Wang, A.; Wang, X.; Wang, J.; Wei, M.-H.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S. C.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.;

Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y.-H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigó, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y.-H.; Coyne, M.; Dahlke, C.; Mays, A. D.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. *"The sequence of the human genome" Science* 2001, *291*, 1304-1351.

(108) Carr, P. W.; Wang, X.; Stoll, D. R. *"Effect of pressure, particle size, and time on optimizing performance in liquid chromatography" Anal Chem* 2009, *81*, 5342-53.

(109) Beaudry, F.; Vachon, P. *"Electrospray ionization suppression, a physical or a chemical phenomenon?" Biomedical Chromatography* 2006, *20*, 200-205.

(110) Walton, D. J.; Lorimer, J. P. "*Polymers*"; Oxford University Press: Oxford ; New York, 2000, p 154 p.

(111) Technologies, C. S. In *Secondary "The Life Expectancy of Solid State Relays"*, February 2011.

(112) Visioli, A. "*Practical PID control*"; Springer: London, 2006, p xviii, 310 p.

(113) Kiker, E. *"Thermocouples versus RTDs" InTech Magazine* 2003.

(114) Saleema, N.; Sarkar, D. K.; Paynter, R. W.; Gallant, D.; Eskandarian, M. *"A simple surface treatment and characterization of AA 6061 aluminum alloy surface for adhesive bonding applications" Appl Surf Sci* 2012, *261*, 742-748.

(115) Zakzeski, J.; Bruijnincx, P. C. A.; Jongerius, A. L.; Weckhuysen, B. M. *"The Catalytic Valorization of Lignin for the Production of Renewable Chemicals" Chem Rev* 2010, *110*, 3552-3599.

(116) Sanderson, K. *"A chewy problem" Nature* 2011, *474*, S12-S14.

(117) Blunk, S. L.; Jenkins, B. M.; Kadam, K. L. *"Combustion properties of lignin residue from lignocellulose fermentation" Biomass: A Growth Opportunity in Green Energy and Value-Added Products, Vols 1 and 2* 1999, 1385-1391.

(118) In *Secondary "Energy Policy Act of 2005"*, 2005.

(119) In *Secondary "Energy Independence and Security Act of 2007"*, 2007.

(120) Mohr, A.; Raman, S. *"Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels" Energ Policy* 2013, *63*, 114-122.

(121) Zhang, Z. B.; Lohr, L.; Escalante, C.; Wetzstein, M. *"Food versus fuel: What do prices tell us?" Energ Policy* 2010, *38*, 445-451.

(122) Lau, M. W.; Gunawan, C.; Dale, B. E. *"The impacts of pretreatment on the fermentability of pretreated lignocellulosic biomass: a comparative evaluation between ammonia fiber expansion and dilute acid pretreatment" Biotechnol Biofuels* 2009, *2*.

(123) Balan, V. *"Current challenges in commercially producing biofuels from lignocellulosic biomass" ISRN Biotechnol* 2014, *2014*, 463074.

(124) Holladay, J.; Bozell, J.; White, J.; Johnson, D. *"Top Value-Added Chemicals from Biomass. Volume II - Results of Screening for Potential Candidates from Biorefinery Lignin" DOE Report PNNL* 2007, *16983*.

(125) Tuck, C. O.; Perez, E.; Horvath, I. T.; Sheldon, R. A.; Poliakoff, M. *"Valorization of Biomass: Deriving More Value from Waste" Science* 2012, *337*, 695-699.

(126) Werpy, T.; Petersen, G.; Aden, A.; Bozell, J.; Holladay, J.; White, J.; Manheim, A.; Eliot, D.; Lasure, L.; Jones, S., "Top value added chemicals from biomass. Volume 1-Results of screening for potential candidates from sugars and synthesis gas"; DTIC Document2004.

(127) Ragauskas, A. J.; Beckham, G. T.; Biddy, M. J.; Chandra, R.; Chen, F.; Davis, M. F.; Davison, B. H.; Dixon, R. A.; Gilna, P.; Keller, M.; Langan, P.; Naskar, A. K.; Saddler, J. N.; Tschaplinski, T. J.; Tuskan, G. A.; Wyman, C. E. *"Lignin Valorization: Improving Lignin Processing in the Biorefinery" Science* 2014, *344*, 709-+.

(128) Pandey, M. P.; Kim, C. S. *"Lignin Depolymerization and Conversion: A Review of Thermochemical Methods" Chem Eng Technol* 2011, *34*, 29-41.

(129) Salvachúa, D.; Karp, E. M.; Nimlos, C. T.; Vardon, D. R.; Beckham, G. T. *"Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria" Green Chemistry* 2015.

(130) Wang, H.; Tucker, M.; Ji, Y. *"Recent development in chemical depolymerization of lignin: a review" Journal of Applied Chemistry* 2013, *2013*.

(131) Rahimi, A.; Azarpira, A.; Kim, H.; Ralph, J.; Stahl, S. S. *"Chemoselective Metal-Free Aerobic Alcohol Oxidation in Lignin" Journal of the American Chemical Society* 2013.

(132) Stahl, S. S.; Coon, J.; Rahimi, A.; Ulbrich, A. In *Secondary "Selective co bond cleavage of oxidized lignin and lignin-type materials into simple aromatic compounds"*; Google Patents, 2014.

(133) Bentley, W. *"The action of nitric acid on vanillin" American Chemical Journal* 1900, *24*, 171 - 181.

(134) Administration, U. S. E. I., "International Energy Outlook 2011": Washington, D.C., 2011.

(135) Agency, U. S. E. P. In *Secondary "RFS2 Data: Spreadsheet of RIN generation and renewable fuel volume production by fuel type for the Renewable Fuel Standard"*, 2015.

(136) U.S. Department of Energy, O. o. S., Office of Biological and Environmental Research, "Lignocellulosic Biomass for Advanced Biofuels

and Bioproducts: Workshop Report, DOE/SC-0170"; 2015.

(137) Palmqvist, E.; Hahn-Hagerdal, B. *"Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification" Bioresource Technol* 2000, *74*, 17-24.

(138) Piotrowski, J. S.; Zhang, Y. P.; Bates, D. M.; Keating, D. H.; Sato, T. K.; Ong, I. M.; Landick, R. *"Death by a thousand cuts: the challenges and diverse landscape of lignocellulosic hydrolysate inhibitors" Front Microbiol* 2014, *5*.

(139) Bellissimi, E.; van Dijken, J. P.; Pronk, J. T.; van Maris, A. J. A. *"Effects of acetic acid on the kinetics of xylose fermentation by an engineered, xylose-isomerase-based Saccharomyces cerevisiae strain" Fems Yeast Res* 2009, *9*, 358-364.

(140) Swinnen, S.; Fernandez-Nino, M.; Gonzalez-Ramos, D.; van Maris, A. J. A.; Nevoigt, E. *"The fraction of cells that resume growth after acetic acid addition is a strain-dependent parameter of acetic acid tolerance in Saccharomyces cerevisiae" Fems Yeast Res* 2014, *14*, 642-653.

(141) Chambel, A.; Viegas, C. A.; Sa-Correia, I. *"Effect of cinnamic acid on the growth and on plasma membrane H+-ATPase activity of Saccharomyces cerevisiae" Int J Food Microbiol* 1999, *50*, 173-179.

(142) Delgenes, J. P.; Moletta, R.; Navarro, J. M. *"Effects of lignocellulose degradation products on ethanol fermentations of glucose and xylose by Saccharomyces cerevisiae, Zymomonas mobilis, Pichia stipitis, and Candida shehatae" Enzyme Microb Tech* 1996, *19*, 220-225.

(143) Iwaki, A.; Ohnuki, S.; Suga, Y.; Izawa, S.; Ohya, Y. *"Vanillin Inhibits Translation and Induces Messenger Ribonucleoprotein (mRNP) Granule Formation in Saccharomyces cerevisiae: Application and Validation of High-Content, Image-Based Profiling" Plos One* 2013, *8*.

(144) Klinke, H. B.; Olsson, L.; Thomsen, A. B.; Ahring, B. K. *"Potential inhibitors from wet oxidation of wheat straw and their effect on ethanol production of Saccharomyces cerevisiae: Wet oxidation and fermentation by yeast" Biotechnol Bioeng* 2003, *81*, 738-747.

(145) Sato, T. K.; Liu, T. J.; Parreiras, L. S.; Williams, D. L.; Wohlbach, D. J.; Bice, B. D.; Ong, I. M.; Breuer, R. J.; Qin, L.; Busalacchi, D.; Deshpande, S.; Daum, C.; Gasch, A. P.; Hodge, D. B. *"Harnessing Genetic Diversity in Saccharomyces cerevisiae for Fermentation of Xylose in Hydrolysates of Alkaline Hydrogen Peroxide-Pretreated Biomass" Appl Environ Microb* 2014, *80*, 540-554.

(146) Zaldivar, J.; Martinez, A.; Ingram, L. O. *"Effect of selected aldehydes on the growth and fermentation of ethanologenic Escherichia coli" Biotechnol Bioeng* 1999, *65*, 24-33.

(147) Larsson, S.; Reimann, A.; Nilvebrant, N. O.; Jonsson, L. J. *"Comparison of different methods for the detoxification of lignocellulose hydrolyzates of spruce" Appl Biochem Biotech* 1999, *77-9*, 91-103.

(148) Jonsson, L. J.; Palmqvist, E.; Nilvebrant, N. O.; Hahn-Hagerdal, B. *"Detoxification of wood hydrolysates with laccase and peroxidase from the white-rot fungus Trametes versicolor" Appl Microbiol Biot* 1998, *49*, 691-697.

(149) Parawira, W.; Tekere, M. *"Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: review"* *Crit Rev Biotechnol* 2011, *31*, 20-31.

(150) Harwood, C. S.; Gibson, J. *"Anaerobic and Aerobic Metabolism of Diverse Aromatic-Compounds by the Photosynthetic Bacterium Rhodopseudomonas-Palustris"* *Appl Environ Microb* 1988, *54*, 712-717.

(151) Schwalbach, M. S.; Keating, D. H.; Tremaine, M.; Marner, W. D.; Zhang, Y. P.; Bothfeld, W.; Higbee, A.; Grass, J. A.; Cotten, C.; Reed, J. L.; Sousa, L. D.; Jin, M. J.; Balan, V.; Ellinger, J.; Dale, B.; Kiley, P. J.; Landick, R. *"Complex Physiology and Compound Stress Responses during Fermentation of Alkali-Pretreated Corn Stover Hydrolysate by an Escherichia coli Ethanologen"* *Appl Environ Microb* 2012, *78*, 3442-3457.

(152) Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; Tabita, F. R.; Gibson, J. L.; Hanson, T. E.; Bobst, C.; Torres, J. L. T. Y.; Peres, C.; Harrison, F. H.; Gibson, J.; Harwood, C. S. *"Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris"* *Nat Biotechnol* 2004, *22*, 55-61.

(153) Egland, P. G.; Pelletier, D. A.; Dispensa, M.; Gibson, J.; Harwood, C. S. *"A cluster of bacterial genes for anaerobic benzene ring biodegradation"* *Proc Natl Acad Sci U S A* 1997, *94*, 6484-6489.

(154) Gibson, J.; Dispensa, M.; Harwood, C. S. *"4-hydroxybenzoyl coenzyme A reductase (dehydroxylating) is required for anaerobic degradation of 4-hydroxybenzoate by Rhodopseudomonas palustris and shares features with molybdenum-containing hydroxylases"* *J Bacteriol* 1997, *179*, 634-642.

(155) Kontur, W. S.; Schackwitz, W. S.; Ivanova, N.; Martin, J.; LaButti, K.; Deshpande, S.; Tice, H. N.; Pennacchio, C.; Sodergren, E.; Weinstock, G. M.; Noguera, D. R.; Donohue, T. J. *"Revised Sequence and Annotation of the Rhodobacter sphaeroides 2.4.1 Genome"* *J Bacteriol* 2012, *194*, 7016-7017.

(156) Mackenzie, C.; Choudhary, M.; Larimer, F. W.; Predki, P. F.; Stilwagen, S.; Armitage, J. P.; Barber, R. D.; Donohue, T. J.; Hosler, J. P.; Newman, J. E.; Shapleigh, J. P.; Sockett, R. E.; Zeilstra-Ryalls, J.; Kaplan, S. *"The home stretch, a first analysis of the nearly completed genome of Rhodobacter sphaeroides 2.4.1"* *Photosynth Res* 2001, *70*, 19-41.

(157) Kim, M. K.; Harwood, C. S. *"Regulation of Benzoate-Coa Ligase in Rhodopseudomonas-Palustris"* *Fems Microbiol Lett* 1991, *83*, 199-203.

(158) Sistrom, W. R. *"The kinetics of the synthesis of photopigments in Rhodopseudomonas spheroides"* *J Gen Microbiol* 1962, *28*, 607-16.

(159) Barbosa, M. J.; Rocha, J. M. S.; Tramper, J.; Wijffels, R. H. *"Acetate as a carbon source for hydrogen production by photosynthetic bacteria"* *J Biotechnol* 2001, *85*, 25-33.

(160) Dutton, P. L.; Evans, W. C. *"The metabolism of aromatic compounds by Rhodopseudomonas palustris. A new, reductive, method of aromatic ring metabolism"* *Biochem J* 1969, *113*, 525-36.

(161) Harwood, C. S.; Burchhardt, G.; Herrmann, H.; Fuchs, G. *"Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway"* *Fems Microbiol Rev* 1998, *22*, 439-458.

(162) Harrison, F. H.; Harwood, C. S. *"The pimFABCDE operon from Rhodopseudomonas palustris mediates dicarboxylic acid degradation and participates in anaerobic benzoate degradation" Microbiol-Sgm* 2005, *151*, 727-736.

(163) Gall, D. L.; Ralph, J.; Donohue, T. J.; Noguera, D. R. *"Benzoyl Coenzyme A Pathway-Mediated Metabolism of meta-Hydroxy-Aromatic Acids in Rhodopseudomonas palustris" J Bacteriol* 2013, *195*, 4112-4120.

(164) Hirakawa, H.; Schaefer, A. L.; Greenberg, E. P.; Harwood, C. S. *"Anaerobic p-Coumarate Degradation by Rhodopseudomonas palustris and Identification of CouR, a MarR Repressor Protein That Binds p-Coumaroyl Coenzyme A" J Bacteriol* 2012, *194*, 1960-1967.

(165) Pan, C.; Oda, Y.; Lankford, P. K.; Zhang, B.; Samatova, N. F.; Pelletier, D. A.; Harwood, C. S.; Hettich, R. L. *"Characterization of anaerobic catabolism of p-coumarate in Rhodopseudomonas palustris by integrating transcriptomics and quantitative proteomics" Molecular & Cellular Proteomics* 2008, *7*, 938-948.

(166) Phattarasukol, S.; Radey, M. C.; Lappala, C. R.; Oda, Y.; Hirakawa, H.; Brittnacher, M. J.; Harwood, C. S. *"Identification of a p-Coumarate Degradation Regulon in Rhodopseudomonas palustris by Xpression, an Integrated Tool for Prokaryotic RNA-Seq Data Processing" Appl Environ Microb* 2012, *78*, 6812-6818.

(167) Larsson, S.; Quintana-Sainz, A.; Reimann, A.; Nilvebrant, N. O.; Jonsson, L. J. *"Influence of lignocellulose-derived aromatic compounds on oxygen-limited growth and ethanolic fermentation by Saccharomyces cerevisiae" Appl Biochem Biotech* 2000, *84-6*, 617-632.

(168) Gu, H. Q.; Zhang, J.; Bao, J. *"Inhibitor analysis and adaptive evolution of Saccharomyces cerevisiae for simultaneous saccharification and ethanol fermentation from industrial waste corncob residues" Bioresource Technol* 2014, *157*, 6-13.

(169) Verduyn, C.; Postma, E.; Scheffers, W. A.; Vandijken, J. P. *"Effect of Benzoic-Acid on Metabolic Fluxes in Yeasts - a Continuous-Culture Study on the Regulation of Respiration and Alcoholic Fermentation" Yeast* 1992, *8*, 501-517.

(170) Fitzgerald, D. J.; Stratford, M.; Gasson, M. J.; Ueckert, J.; Bos, A.; Narbad, A. *"Mode of antimicrobial action of vanillin against Escherichia coli, Lactobacillus plantarum and Listeria innocua" J Appl Microbiol* 2004, *97*, 104-113.

(171) Lemke, R. A. S.; Peterson, A. C.; Ziegelhoffer, E. C.; Westphall, M. S.; Tjellstrom, H.; Coon, J. J.; Donohue, T. J. *"Synthesis and scavenging role of furan fatty acids" Proc Natl Acad Sci U S A* 2014, *111*, E3450-E3457.

(172) Lemmer, K. C.; Dohnalkova, A. C.; Noguera, D. R.; Donohue, T. J. *"Oxygen-Dependent Regulation of Bacterial Lipid Production" J Bacteriol* 2015, *197*, 1649-1658.

(173) Austin, S. *Utilization of corn stover hydrolysates by Rhodobacter sphareoides and Rhodopseudomonas palustris under photoheterotrophic conditions*. University of Wisconsin-Madison, Madison, WI, 2013.

(174) Roe, A. J.; O'Byrne, C.; McLaggan, D.; Booth, I. R. *"Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity" Microbiol-Sgm* 2002, *148*, 2215-2222.

(175) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *"The path forward for biofuels and biomaterials" Science* 2006, *311*, 484-489.

(176) Ralph, J.; Brunow, G.; Boerjan, W. In *eLS*; John Wiley & Sons, Ltd, 2001.

(177) Lin, S. Y.; Lin, I. S. "*Ullmann's encyclopedia of industrial chemistry"*, 5th, completely rev. ed.; VCH: Weinheim, Federal Republic of Germany ; Deerfield Beach, FL, USA, 1985, p A15.

(178) Borregaard. In *Secondary "Business Areas"*.

(179) Hocking, M. B. *"Vanillin: Synthetic flavoring from spent sulfite liquor" J Chem Educ* 1997, *74*, 1055-1059.

(180) Gierer, J.; Noren, I. *"Oxidative Pretreatment of Pine Wood to Facilitate Delignification during Kraft Pulping" Holzforschung* 1982, *36*, 123-130.

(181) Ljunggren, S.; Olsson, A. *"The Specificity in Oxidation of Some Lignin and Carbohydrate Models and Pine Wood Shavings with Permanganate and Pyridinium Dichromate before the Kraft Pulping Process" Holzforschung* 1984, *38*, 91-99.

(182) Argyropoulos, D. S.; Sun, Y. J. *"Photochemically induced solid-state degradation, condensation, and rearrangement reactions in lignin model compounds and milled wood lignin" Photochem Photobiol* 1996, *64*, 510-517.

(183) Chatel, G.; Rogers, R. D. *"Review: Oxidation of Lignin Using Ionic Liquids-An Innovative Strategy To Produce Renewable Chemicals" Acs Sustain Chem Eng* 2014, *2*, 322-339.

(184) Collinson, S. R.; Thielemans, W. *"The catalytic oxidation of biomass to new materials focusing on starch, cellulose and lignin" Coordin Chem Rev* 2010, *254*, 1854-1870.

(185) Crestini, C.; DAuria, M. *"Singlet oxygen in the photodegradation of lignin models" Tetrahedron* 1997, *53*, 7877-7888.

(186) Fukagawa, N.; Ishizu, A. *"Photoreaction of Phenacyl Aryl Ether Type Lignols" J Wood Chem Technol* 1991, *11*, 263-289.

(187) Nguyen, J. D.; Matsuura, B. S.; Stephenson, C. R. J. *"A Photochemical Strategy for Lignin Degradation at Room Temperature" Journal of the American Chemical Society* 2014, *136*, 1218-1221.

(188) Vanucci, C.; Deviolet, P. F.; Bouaslaurent, H.; Castellan, A. *"Photodegradation of Lignin - a Photophysical and Photochemical Study of a Non-Phenolic Alpha-Carbonyl Beta-O-4 Lignin Model Dimer, 3,4-Dimethoxy-Alpha-(2'-Methoxyphenoxy)Acetophenone" J Photoch Photobio A* 1988, *41*, 251-265.

(189) Chang, H. M.; Cowling, E. B.; Brown, W.; Adler, E.; Miksche, G. *"Comparative Studies on Cellulolytic Enzyme Lignin and Milled Wood Lignin of Sweetgum and Spruce" Holzforschung* 1975, *29*, 153-159.

(190) Ralph, J.; Landucci, L. L. "*Lignin and lignans : advances in chemistry*"; CRC Press: Boca Raton, 2010, p xvi, 651 p.

(191) Fouad, F. M.; Farrell, P. G. *"Primary and Secondary Kinetic Isotope Effects in E2-Elimination Reactions" Tetrahedron Lett* 1978, 4735-4738.

(192) Toledano, A.; Serrano, L.; Balu, A. M.; Luque, R.; Pineda, A.; Labidi, J. *"Fractionation of Organosolv Lignin from Olive Tree Clippings and its Valorization to Simple Phenolic Compounds" Chemsuschem* 2013, *6*, 529-536.

(193) Xu, W. Y.; Miller, S. J.; Agrawal, P. K.; Jones, C. W. *"Depolymerization and Hydrodeoxygenation of Switchgrass Lignin with Formic Acid" Chemsuschem* 2012, *5*, 667-675.

(194) Bauer, K.; Garber, D.; Surburg, H. "*Ullmann's encyclopedia of industrial chemistry*", 5th, completely rev. ed.; VCH: Weinheim, Federal Republic of Germany ; Deerfield Beach, FL, USA, 1985.

(195) Bjorsvik, H. R.; Liguori, L. *"Organic processes to pharmaceutical chemicals based on fine chemicals from lignosulfonates" Org Process Res Dev* 2002, *6*, 279-290.

(196) Martinez, A. T.; Rencoret, J.; Marques, G.; Gutierrez, A.; Ibarra, D.; Jimenez-Barbero, J.; del Rio, J. C. *"Monolignol acylation and lignin structure in some nonwoody plants: A 2D NMR study" Phytochemistry* 2008, *69*, 2831-2843.

(197) Luterbacher, J. S.; Rand, J. M.; Alonso, D. M.; Han, J.; Youngquist, J. T.; Maravelias, C. T.; Pfleger, B. F.; Dumesic, J. A. *"Nonenzymatic Sugar Production from Biomass Using Biomass-Derived gamma-Valerolactone" Science* 2014, *343*, 277-280.

(198) Grimsrud, P. A.; Carson, J. J.; Hebert, A. S.; Hubler, S. L.; Niemi, N. M.; Bailey, D. J.; Jochem, A.; Stapleton, D. S.; Keller, M. P.; Westphall, M. S.; Yandell, B. S.; Attie, A. D.; Coon, J. J.; Pagliarini, D. J. *"A quantitative map of the liver mitochondrial phosphoproteome reveals posttranslational control of ketogenesis" Cell Metab* 2012, *16*, 672-83.

(199) Still, A. J.; Floyd, B. J.; Hebert, A. S.; Bingman, C. A.; Carson, J. J.; Gunderson, D. R.; Dolan, B. K.; Grimsrud, P. A.; Dittenhafer-Reed, K. E.; Stapleton, D. S.; Keller, M. P.; Westphall, M. S.; Denu, J. M.; Attie, A. D.; Coon, J. J.; Pagliarini, D. J. *"Quantification of Mitochondrial Acetylation Dynamics Highlights Prominent Sites of Metabolic Regulation" J Biol Chem* 2013, *288*, 26209-26219.

(200) Pagliarini, D. J.; Calvo, S. E.; Chang, B.; Sheth, S. A.; Vafai, S. B.; Ong, S. E.; Walford, G. A.; Sugiana, C.; Boneh, A.; Chen, W. K.; Hill, D. E.; Vidal, M.; Evans, J. G.; Thorburn, D. R.; Carr, S. A.; Mootha, V. K. *"A mitochondrial protein compendium elucidates complex I disease biology" Cell* 2008, *134*, 112-123.

(201) Crane, F. L. *"The Discovery of Coenzyme-Q - a Commentary on Isolation of a Quinone from Beef-Heart Mitochondria" Biochim Biophys Acta* 1989, *1000*, 358-363.

(202) Quinzii, C. M.; DiMauro, S.; Hirano, M. *"Human coenzyme Q10 deficiency" Neurochem Res* 2007, *32*, 723-7.

(203) Emmanuele, V.; Lopez, L. C.; Berardo, A.; Naini, A.; Tadesse, S.; Wen, B.; D'Agostino, E.; Solomon, M.; DiMauro, S.; Quinzii, C.; Hirano, M. *"Heterogeneity of coenzyme Q10 deficiency: patient study and literature review" Arch Neurol* 2012, *69*, 978-83.

(204) He, C. W. H.; Xie, L. T. X.; Allan, C. M.; Tran, U. C.; Clarke, C. F. *"Coenzyme Q supplementation or over-expression of the yeast Coq8 putative kinase stabilizes multi-subunit Coq polypeptide complexes in yeast coq null mutants" Bba-Mol Cell Biol L* 2014, *1841*, 630-644.

(205) Ulbrich, A.; Minogue, C. E.; Stefely, J. A.; Lohman, D. C.; Reidenbach, A. G.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. In *Secondary "Mass Spectrometric Characterization of Coenzyme Q Biosynthesis"*: St. Louis, MO, 2015.

(206) Leonard, C. J.; Aravind, L.; Koonin, E. V. *"Novel families of putative protein kinases in bacteria and archaea: Evolution of the "eukaryotic" protein kinase superfamily" Genome Res* 1998, *8*, 1038-1047.

(207) Horvath, P. O.; Czermin, B.; Gulati, S.; Pyle, A.; Hassani, A.; Foley, C.; Taylor, R. W.; Chinnery, P. F. *"Adult-Onset Cerebellar Ataxia Due to Mutations in the Cabc1/Adck3 Gene" J Neurol Neurosur Ps* 2012, *83*.

(208) Marbois, B.; Xie, L. X.; Choi, S.; Hirano, K.; Hyman, K.; Clarke, C. F. *"para-Aminobenzoic acid is a precursor in coenzyme Q6 biosynthesis in Saccharomyces cerevisiae" J Biol Chem* 2010, *285*, 27827-27838.

(209) Pierrel, F.; Hamelin, O.; Douki, T.; Kieffer-Jaquinod, S.; Mühlenhoff, U.; Ozeir, M.; Lill, R.; Fontecave, M. *"Involvement of mitochondrial ferredoxin and para-aminobenzoic acid in yeast coenzyme Q biosynthesis" Chemistry & biology* 2010, *17*, 449-459.